任永安　主编



社区矫正信息系统工程设计与安全

THE ENGINEERING DESIGN AND SECURITY OF INFORMATION SYSTEM FOR THE COMMUNITY CORRECTION

清华大学出版社

社区矫正信息系统工程设计与安全

任永安　主编
李军怀　副主编
张　璟　张亚玲　编著

清华大学出版社
北　京

内 容 简 介

社区矫正工作是党中央、国务院在新形势下做出的重要战略部署，是创新社会管理、维护社会和谐稳定的重要举措，是司法体制和工作机制改革的重要内容。在社区矫正管理工作中实施信息化高科技支撑，加强信息化建设，是实施科技强警战略的迫切需要，具有非常重要的现实意义。

本书以社区矫正信息化建设规划和实施为目标，全面介绍了社区矫正信息化建设中涉及的新思路、新技术和新方法。全书共分 6 章，全面系统地介绍了物联网、云计算、大数据等社区矫正信息系统建设支撑技术，并重点介绍了社区矫正信息系统体系结构、数据集成技术、安全策略及系统设计与实现的方法。

本书可供司法行政机关、社区矫正机构及相关行业的管理与技术人员作为业务用书，也可作为司法、警察、社工等专业教材或参考书使用。

图书在版编目(CIP)数据

社区矫正信息系统工程设计与安全/任永安主编. --北京：清华大学出版社，2015
ISBN 978-7-302-39039-8

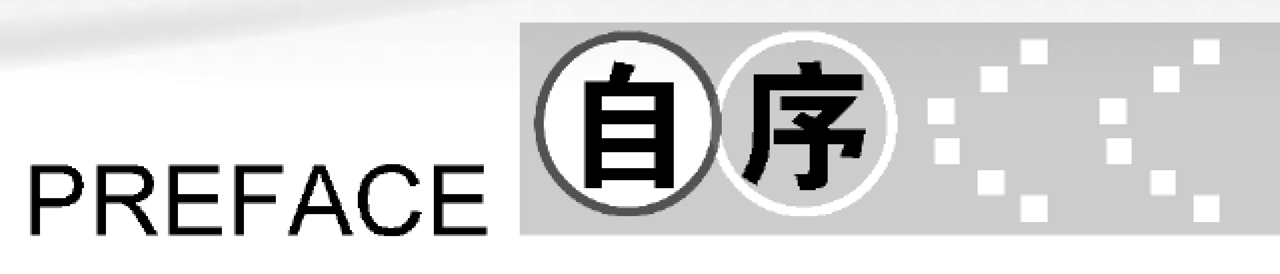
Ⅰ. ①社…　Ⅱ. ①任…　Ⅲ. ①社区—监督改造—管理信息系统—系统工程—中国
Ⅳ. ①D926.7-39

中国版本图书馆 CIP 数据核字(2015)第 017121 号

责任编辑：刘向威　李　晔
封面设计：文　静
责任校对：李建庄
责任印制：

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　**邮　　编**：100084
社 总 机：010-62770175　　**邮　　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：http://www.tup.com.cn,010-62795954
印 刷 者：
装 订 者：
经　　销：全国新华书店
开　　本：170mm×230mm　　**印　　张**：15.75　　**字　　数**：347 千字
版　　次：2015 年 4 月第 1 版　　**印　　次**：2015 年 4 月第1 次印刷
印　　数：1～　000
定　　价：　.00 元

产品编号：061695-01



2013年9月，在司法部社区矫正管理局姜爱东局长、司法部司法研究所罗厚如所长的大力支持下，我联合西安理工大学计算机专家张璟教授、河北经贸大学法学院的法学专家王利军教授，与司法部社区矫正管理局的实务管理人员、司法部司法研究所和河北经贸大学的法律研究人员以及西安理工大学计算机科学与工程学院的技术专家为团队成员，成立了“社区矫正信息化建设战略问题研究”项目组，撰写项目研究计划，申报科技部2013年度国家软科学重大合作项目，经过初评、专题答辩等程序，获得评审专家认可，同年12月，科技部正式批准立项。

“社区矫正信息化建设战略问题研究”是“文理交叉”的战略性应用对策研究，目的在于为实施社区矫正信息化战略决策及其信息系统工程设计提供法律依据和技术参考，创新性比较强。项目组由我总负责，姜爱东局长为顾问，张璟教授主要负责技术部分。为充分发挥项目团队各个成员的专业特长，使项目成果丰富并具有前瞻性，经协商，决定把项目下分两个子课题开展研究，分别为法律篇和技术篇，研究团队也相应地分为法律组和技术组。法律组主要研究、论述实行社区矫正制度的法律规定及其社区矫正信息化建设的法律依据；技术组主要研发、介绍社区矫正信息系统工程设计思路、方法和安全防护策略。我除了负责整个项目的研究工作外，还和王利军教授共同负责法律组的研究工作；张璟教授和李军怀教授负责技术组的研究工作。本书即是项目技术篇的最终成果。

我对社区矫正工作及其制度的关注与兴趣由来已久。2003年7月，经中央批准，司法部联合最高人民法院、最高人民检察院、公安部印发了《关于开展社区矫正试点工作的通知》，确定在北京、江苏等6省市开展社区矫正试点工作。同年11月我开始负责司法部部级课题管理工作，这既包括面向法学界、政法实务部门设立的“国家法治建设与法学理论研究项目”(以下简称“社会课题”)，又包括对全国司法行政系统单位设立的“全国司法行政系统理论研究规划项目”(以下简称“系统课

题”)。在同年申报的社会课题和系统课题中就出现了多个关于社区矫正研究的申请书,这在当时还是比较新颖的选题。社会课题是司法部出资设立的,对于司法行政工作研究的课题,评审专家格外关注,而对社区矫正问题的研究更是如此。当年,专家组推荐了两项社区矫正课题立项,分别为中国政法大学王平教授主持的重点课题和司法部刘和平局长主持的委托课题。随后,在当年度的系统课题中,列入了包括司法部副部长胡泽君主持的重大课题在内的两项关于社区矫正的课题。由此,我对社区矫正开始关注并逐渐产生兴趣,对其在我国刑事司法方面的重大变革和增强司法行政职责功能方面的重大作用,认识上更加深刻而具体。此后在我负责司法部部级课题管理工作的10年内,每年在制定年度课题指南目录时,必定要包含社区矫正研究的课题,从而引导法学理论界的专家学者和政法实务部门的相关领导对社区矫正相关问题组织研究。在我负责司法部部级课题管理期间,每年立项的社会课题和系统课题中都有关于社区矫正的项目。现今,看到很多带有“司法部部级科研项目”标识的社区矫正成果出版物、注明“属于司法部课题成果”的社区矫正理论文章,特别是看到我国社区矫正事业蓬勃发展,我为多年来能够支持社区矫正理论研究,引导专家学者为社区矫正制度的实施与完善提供更多智力成果而甚感欣慰!

在刑事司法领域,对于法院作出判决有罪的刑事犯罪人员(简称罪犯),对其惩罚和改造主要有两种模式:集中在监狱中服刑和分散于社区服刑。前者为监狱矫正,是监禁刑罚执行制度;后者为社区矫正,是非监禁刑罚执行制度。其中,社区矫正制度在我国实行的时间不长,从2003年在6个省市开展试点工作开始算起,经过2005年在河北、内蒙古等12个省、自治区、直辖市扩大试点,2009年在全国全面试行,直至当前,虽然只有11年的时间,但发展迅速,成效显著,承担了数量巨大的罪犯教育改造任务,为贯彻宽严相济的刑事政策,保持社会和谐稳定,服务经济社会发展作出了突出贡献。然而,社区矫正制度在我国来说,毕竟是一项新鲜事物,在其试行直至全面实施的运行过程中,存在很多问题和不足,这既有工作机制方面的问题,也有制度建设方面的问题,需要专家学者深入社区矫正实际,认真研究分析,对改进和完善我国的社区矫正制度提出意见和建议。

社区矫正工作是司法行政系统的一项重要工作,社区矫正制度也是中国特色社会主义司法行政制度的重要组成部分。司法部官方对社区矫正给出了科学定义:“社区矫正是指将管制、缓刑、暂予监外执行、假释、剥夺政治权利并在社会上服刑的符合法定条件的罪犯置于社区内,由专门的国家机关在相关社会团体、民间组织和社会志愿者的协助下,在判决、裁定或决定确定的期限内,矫正其犯罪心理和行为恶习,促进其顺利回归社会的非监禁刑罚执行活动。”

我国对于罪犯的教育改造，传统理念及做法是在监狱中集中管理，而将符合一定条件的罪犯放入社会进行监督管理和教育矫正，这种全新的教育改造方式，既需要社区人员转变观念，愿意接受，能够与社区服刑人员和平共处；还要转变政府管理人员旧的思想观念和教育改造手段，能够实现法律刑罚效果，维护社会正常秩序。

鉴于社区矫正工作的极端重要性和罪犯身份的社会敏感性，党中央、国务院对于该项工作向来高度重视，2013 年 11 月党的十八届三中全会通过的《中共中央关于全面深化改革若干重大问题的决定》明确提出，要“健全社区矫正制度”。2014 年 4 月 21 日，习近平总书记在听取司法部工作汇报时再次作出重要指示，明确指出：“社区矫正已在试点的基础上全面推开，新情况新问题会不断出现。要持续跟踪完善社区矫正制度，加快推进立法，理顺工作体制机制，加强矫正机构和队伍建设，切实提高社区矫正工作水平。”

社区矫正制度在我国发展速度之快，在社区中服刑的人员数量增幅之大，远远超出包括我本人在内的诸多关注这项制度的人的预计。总的来看，我国社区矫正工作取得的巨大成就有目共睹，主要包括如下三个方面。一是社区矫正人员的数量大幅增加，社区矫正工作人员完成了重要教育改造任务。截至 2014 年 8 月份，全国在册社区服刑人员由 2009 年的 21.5 万人增加到 70.9 万人，各地累计接收社区服刑人员 184.7 万人，累计解除社区服刑人员 113.8 万人，社区服刑人员在矫正期间重新犯罪率一直处在 0.2%的较低水平，取得了良好的法律效果和社会效果。二是在做好社区矫正人员监管工作的同时，司法部积极争取中央领导和相关部门支持，及时出台和修改相关法律规定，从制度建设和政策完善方面加强工作力度。2011 年《刑法修正案（八）》和 2012 年《刑事诉讼法修订》对社区矫正制度作出了明确规定，标志着我国社区矫正法律制度正式确立；2012 年 1 月，最高人民法院、最高人民检察院、公安部、司法部联合制定出台了《社区矫正实施办法》，将各地在实践中形成的行之有效的工作体制机制、矫正方法和模式等固定下来，上升为统一的制度，使之成为社区矫正工作的操作规范和基本依据，为社区矫正工作提供了制度保障。司法部还下发了《关于印发和使用＜社区矫正执法文书格式＞的通知》等配套规定。此外，各省市司法厅局根据本地区实际情况，相继制定《实施细则》，社区矫正法律体系逐渐确立。目前，司法部正在推动社区矫正国家立法。三是始终重视现代化高科技技术在社区矫正工作中的运用，已经初步建立了“全国社区矫正信息管理系统”，出台了《社区矫正管理信息系统技术规范》和《社区矫正人员定位系统技术规范》等规范性技术标准。

我国社区矫正制度发展迅速，但在实际工作中，由于管理人员人手少、经验不

足，出现管理困难，如不借助现代高科技手段，将难以承担对逐年增加的社区服刑罪犯进行教育矫正的重任。为此，社区矫正信息化建设显得非常必要和紧迫。2014年5月27日，中共中央政治局委员、中央政法委书记孟建柱在全国社区矫正工作会议上对做好社区矫正信息化工作，明确提出："要充分利用现代科技手段，提高社区矫正现代化水平。现代科技手段的发展，已成为促进经济社会发展的强大力量，也为社区矫正提供了新的平台和手段。特别是大数据时代已经来临，谁率先拥有大数据、善于利用大数据，谁就能掌握主动、赢得未来。各地各有关部门要总结推广一些地方探索的'网上枫桥经验'，通过开办网上'诊所'等方式，及时了解并帮助解决社区服刑人员的思想、心理等问题，充分发挥信息网络在教育矫正社区服刑人员方面的积极作用。要树立基础工作信息化的理念，确保对社区服刑人员各类基础信息做到及时有效掌握，提高社区矫正工作信息化、现代化水平。要把传统有效做法与现代科技手段有机结合起来，及时发现、提醒超越活动范围的社区服刑人员，提高依法监管工作水平。"

现代高科技信息技术突飞猛进，全球已进入信息化时代，物联网、云计算、大数据等信息技术新概念、新模式、新成果，不断涌现，层出不穷。高新信息技术在国家治理、社会秩序、群众生活、自然生态等方面的运用越来越迅速、普及、深入，其功能作用在各个领域得到充分发挥和体现。与此同时，近年来，随着我国经济社会的不断发展，更多新信息技术被引入政府管理活动中，为科学高效、规范有序的政府管理提供优质服务。为此，我国社区矫正工作自开始创立，就高度重视信息化技术的建设与使用，依托信息化技术手段，增强社区矫正管理能力和水平，弥补管理方面的不足，健全社区矫正管理的手段和方式。

在社区矫正工作中使用现代高科技信息技术，就要研究社区矫正与信息技术的完美结合，探索、寻找两者的客观规律和自身特性，特别是熟练掌握社区矫正全程运行细节，将信息技术融合到社区矫正管理的各个阶段，开发相应的信息管理模块，再整合、兼容各个管理模块，最终形成完整的社区矫正信息技术系统。此外，还要研究制定相应的技术规范、技术标准，以及系统运行的安全防范技术；当然，对于技术成本、技术前瞻、实用效能、网络兼容等方面也要进行充分考虑和事前论证。

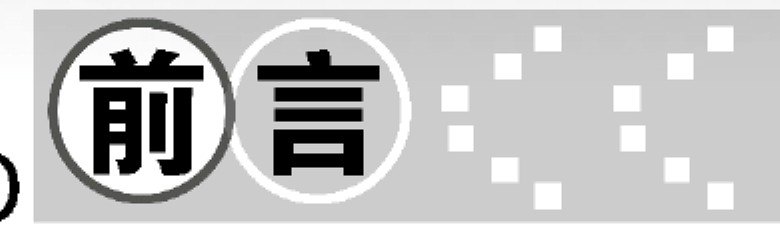
本项目课题组下的法律组和技术组成员，分别从社区矫正的法律问题和技术问题两个方面开展论述和编著工作。法律组就社区矫正的概念、理论渊源、法律制度、实践问题、立法、国外经验，以及社区矫正信息化的法律问题和法律制度构建等，进行详细论述；技术组先就社区矫正涉及的物联网、云计算、大数据等信息支撑技术做了全面系统的分析，再重点研究社区矫正信息系统体系结构、数据集成策略、安全策略及系统设计与实现的方法。因两组成员的专业背景不同，对社区矫正

信息技术的研究角度和研究方法也就不同，最终成果内容不同，各具特色，其中，技术组的成果已先行成书，法律组的成果将后续出版。

我作为一名法律专业人员，带领计算机技术专家和其他法学专业人员，组成研究团队，对社区矫正信息化的法律问题和技术问题进行专题研究，这本身就是一种大胆创新，至于研究成果难免出现瑕疵，敬请各位专家学者、社区矫正工作者批评赐教，共同致力于为发展和完善具有中国特色的社会主义社区矫正制度而努力！

任永安

2014 年 9 月

FOREWORD 

前言

社区矫正(community correction)是一种不使罪犯与社会隔离并利用社区资源教育改造罪犯的方法,是所有在社区环境中管理教育罪犯方式的总称。国外较常见的适用社区矫正的方式包括缓刑、假释、社区服务、暂时释放、中途之家、工作释放、学习释放等。在我国,社区矫正是指将符合社区矫正条件的罪犯置于社区内,由专门的国家机关,在相关社会团体和民间组织以及社会志愿者的协助下,在判决、裁定或决定确定的期限内,矫正其犯罪心理和行为恶习,并促进其顺利回归社会的非监禁刑罚执行活动。

随着我国维护社会稳定任务的日益艰巨,各级政府非常重视社区矫正工作。而社区矫正信息化建设是当前各级司法行政管理部门必须加强重要工作之一。司法部部长吴爱英多次提出:要加强社区矫正信息化建设,建立完善社区矫正工作信息平台、社区矫正人员信息库、社区矫正工作信息技术规范等工作,不断提高社区矫正信息化应用水平;要努力提高现代信息技术在社区服刑人员等特殊人群管理中的应用水平,依靠科技进步,推进对社区服刑人员等特殊人群的管理创新。

我国社区矫正信息化建设的重要性主要包括以下三方面:

其一,社区矫正的信息化建设,是社会管理创新在新形势下的具体体现。社区矫正是社会管理的一项重要职能,应遵循社会的发展规律,根据政治、经济和社会的发展趋势,不断地改进管理观念、体制和手段,提高社区矫正的能力和水平。随着社会信息化程度的提高,高科技与人们生活和工作的关系也越来越密切,社区矫正的管理也从静态封闭式的环境走向了动态开放式的环境,社会服刑人员的控制难度随之增大。只有不断地改进社区矫正工作手段,加强社区矫正的信息化建设的信息含量和发展速度,才能使社区矫正工作与社会发展、社会管理的需求同步。

其二,社区矫正信息化建设是发展的迫切需要。社区矫正工作经过这些年的

探索，形成了具有中国特色的体系和结构，前景良好。但现有的工作手段和资源已经不能适应社区矫正工作的需求，服刑人员人数与工作人员人数比例严重失衡，社区矫正工作中监管、教育的动态管理很难真正实现，资料衔接上的缺口和漏管等问题亟须借助现代化科技手段来解决。由于社区矫正工作具有开放性、动态性和即时性，只有依靠信息化和现代化才能充分整合资源，建立一个全新快速的工作模式。传统的人工管理方法几乎无法做到全面的监管，也无法进行全面的管理。电子监控设备的加入，使工作人员的工作效率大大提高，也降低了人工管理的成本。

其三，社区矫正信息化建设是适应国际行刑发展潮流的迫切需要。由于西方发达国家社区矫正工作起步早，其管理制度随之比较成熟，工作方式和管理手段也相对比较完备，科技化含量高，社区矫正效果好。但是，我国的社区矫正不能照抄全搬其他国家的模式，应该因地制宜，结合实际国情和社会发展现状来提高科技水平，探索一条适合我国发展的社区矫正信息化建设的道路。

本书在技术层面围绕社区矫正信息化技术与方法进行了论述，全书分为6章。第1章介绍了物联网的基础技术，包括物联网编码、射频识别技术、生物识别技术、位置服务技术，及物联网在基层社区矫正信息系统中的应用；第2章首先介绍了云计算的基本概念、特征、分类以及大数据技术的概念与关键技术，然后介绍云计算平台体系结构、关键技术、大数据处理技术，以及典型云计算支撑环境与工具、云计算应用迁移与部署技术，最后介绍了云计算和大数据在省市级社区矫正信息系统中的应用思路；第3章介绍了基于云计算与物联网技术的社区矫正信息系统总体架构、系统软件结构、网络支撑结构以及相关关键技术；第4章首先分析了社区矫正信息系统数据集成的需求，然后介绍了信息系统数据集成的基本模式、策略，最后给出了基于SOA/ESB的社区矫正信息系统数据集成方案；第5章介绍了社区矫正系统信息安全技术基础、安全策略、安全框架，以及社区矫正信息系统构建中的主机安全技术和云计算安全技术；第6章首先介绍了面向省市的社区矫正信息系统建设的基本原则、工作流程、系统逻辑结构，然后介绍了省市级社区矫正信息系统中心的网络连接结构及中心网络结构，最后介绍了基层社区矫正管理软件系统的详细设计与实现过程。

本书由司法部司法研究所任永安研究员、西安理工大学计算机科学与工程学院张璟教授和李军怀教授共同策划编写。其中，第1、2、3、6章由李军怀编写，第4章由张璟编写，第5章由张亚玲编写；任永安负责全书统稿编校工作并重点修改完善书稿中的法律内容。

本书参考和引用了大量的相关成果，书后已一一列出相应的参考文献及作者。本书也是科技部2013年国家软科学重大合作项目“社区矫正信息化建设战略研究”(批准编号2013GXS2B010)的最终成果之一，得到科技部专项经费资助。在项

目研究和书稿编写过程中，得到司法部办公厅科技办吴键处长、科技部办公厅副主任兼调研室主任胥和平研究员，以及司法部社区矫正管理局、司法部司法研究所、西安理工大学、河北经贸大学等相关部门和领导的大力支持。诚然，书稿能够正式出版，得益于清华大学出版社及刘向威博士的鼎力支持。在此，谨向各位表示诚挚敬意和衷心感谢！

编者

2014 年 9 月


CONTENTS
目录



物联网基础

物联网技术的迅速发展和广泛应用,为社区矫正信息化提供了重要的技术支撑。通过物联网技术,可以实现对社区矫正对象的智能识别、定位、跟踪、监控等管理,为建设面向未来的智慧社区矫正系统提供帮助。

1.1 物联网概述

物联网(The Internet of Things)的概念是在1999年提出的,即通过射频识别(Radio Frequency IDentification,RFID)、红外感应器、全球定位系统、激光扫描器等信息传感设备,按约定的协议,把任何物品与互联网连接起来,进行信息交换和通信,以实现智能化识别、定位、跟踪、监控和管理的一种网络。简言之,物联网就是"物物相连的互联网"。物联网通过智能感知、识别技术与普适计算、广泛应用于网络的融合中,也因此被称为继计算机、互联网之后世界信息产业发展的第三次浪潮,被视为互联网的应用拓展,是新一代信息技术的重要组成部分。

物联网至今没有统一的定义,大家众说纷纭。有的认为RFID的互联就是物联网,有的认为传感网络就是物联网,有的认为M2M(Machine to Machine)就是物联网,有的认为把互联网的用户端延伸和扩展到任何物体与物体之间就是物联网。

国际电信联盟(International Telecommunication Union,ITU)对物联网的定义为:物联网实现物到物(Thing to Thing)、人到物(Human to Thing)和人到人(Human to Human)的互连。这里人与物的互连是指使用传感器等设备实现人与物体的互连,而人与人的互连是指使用传感系统而不是计算机实现人与人之间的

互连。物联网的核心是实现物体(包含人)之间的互连,从而能够实现物体与物体之间的信息交换和通信。

欧盟关于物联网的定义是:物联网是未来互联网的一个组成部分,是基于标准和交互通信协议的具有自配置能力的动态全球网络设施,在物联网内物理和虚拟的“物件”具有身份、物理属性、拟人化等特征,它们能够被一个综合的信息网络所连接。

在物联网中,物体信息通过网络传输到信息处理中心后可实现各种信息服务和应用。物联网的主要作用是缩小物理世界和信息系统之间的距离,它可以通过射频识别(RFID)、传感器、全球定位系统等感知、识别、定位等设备,按约定的协议,将物体连接到信息网络中,实现物理空间和信息空间的融合。

从上面的几种概念中,可以看出物联网概念有两层含义:第一,物联网的核心和基础仍然是互联网,是在互联网基础上的延伸和扩展;第二,其用户端延伸和扩展到了任何物体与物体之间,进行信息交换和通信。与传统的互联网相比,物联网的具有如下特征:

(1) 物联网是各种感知技术的融合应用。物联网中部署了海量的各种类型的感知设备,每个感知设备构成一个信息源,不同类别的感知设备形成一个多源多模信息集合。感知设备获得的数据具有实时性,按一定的频率周期性地采集环境信息,不断更新数据。

(2) 物联网是建立在互联网上的泛在网络。物联网技术基础和核心是互联网,通过各种有线和无线网络与互联网融合,将物体的信息实时准确地传递出去。在物联网上的实时信息通过网络传输,形成海量数据,同时必须适应各种异构网络和协议。

(3) 物联网不仅提供了各种感知设备的连接,其本身也具有智能处理的能力,能够对物体实施智能控制。物联网将传感器和智能处理结合起来,利用云计算、模式识别等各种智能技术,扩充其应用领域。从感知设备获得的海量数据中分析、加工和处理出有意义的信息,以适应不同用户的不同需求,发现新的应用领域和应用模式。

1.1.1 物联网发展历程

“物联网”的概念于1999年由麻省理工学院的Auto-ID实验室提出,“万物皆可通过网络互联”,阐明了物联网的基本含义。Auto-ID的物联网概念是以无线传感器网络和射频识别技术为支撑,随着技术和应用的发展,物联网的内涵已经发生了较大变化。

1999年在美国召开的移动计算和网络国际会议Mob-iCom上提出了传感网

(智能尘埃)是 21 世纪人类面临的又一个发展机遇。同年,麻省理工学院的 Gershenfeld Neil 教授撰写了 *When Things Start to Think* 一书,以这些为标志开始了物联网的发展。

2005 年 11 月 17 日,在突尼斯举行的信息社会世界峰会(WSIS)上,国际电信联盟(ITU)发布了《ITU 互联网报告 2005:物联网》,正式提出了“物联网”的概念。报告指出:无所不在的“物联网”通信时代即将来临,世界上所有的物体都可以通过互联网主动进行信息交换。通过一些关键技术,用互联网将世界上的物体都连接在一起,使世界万物都可以上网。这些关键技术包括通信技术、RFID、传感器、机器人技术、嵌入式技术和纳米技术等。

2006 年 3 月,欧盟召开题为“From RFID to the Internet of Things”的会议,对物联网做了进一步的描述,并于 2009 年制定了物联网研究策略的路线图。

2009 年,IBM 首席执行官 Samuel J. Palmisano 提出了“智慧地球”(Smart-Planet)的概念,把传感器嵌入和装备到电网、铁路、桥梁、隧道、公路、建筑、油气管道等各种应用中,并且通过智能处理,达到智慧状态。

物联网被预言为继互联网之后全球信息产业的又一次科技与经济浪潮,受到各国政府、企业和学术界的重视,美国、欧盟、日本等甚至将其纳入国家和区域信息化战略。目前,美国、欧盟、日本和韩国等都在投入巨资深入研究探索物联网。

2004 年日本总务省(MIC)提出 u-Japan 计划,该战略力求实现人与人、物与物、人与物之间的连接,希望将日本建设成一个随时、随地、任何物体、任何人均可连接的泛在网络社会。2009 年 7 月,日本信息技术战略本部发布了《i-Japan 战略 2015》,将目标聚焦在三大公共事业,即电子化政府治理、医疗健康信息服务、教育与人才培育,提出到 2015 年,通过物联网技术达到“新的行政改革”,使行政流程简化、效率化、标准化、透明化,同时推动电子病历、远程医疗、远程教育等应用的发展。

韩国于 2006 年提出 u-Korea 计划,该计划旨在建立无所不在的社会(ubiquitous society),在民众的生活环境里建设智能网络(如 IPv6、BCN①、USN②)和各种新型应用(如 DMB③、Telematics、RFID),让民众可以随时随地享有科技智慧服务。2009 年 10 月,韩国通信委员会出台了《物联网基础设施构建基本规划》,确定了通过实施“构建物联网基础设施、发展物联网服务、研发物联网技术、营造物联网推广环境”4 大领域相关的 12 项课题,提出到 2012 年实现“通过构建世界最先

① 宽带融合网络:Broadband convergence Network。
② 无所不在的感测网路:Ubiquitous Sensor Network。
③ 数字多媒体广播:Digital Multimedia Broadcasting。

进的物联网基础实施，打造未来广播通信融合领域超一流信息通信技术强国”的目标。

2009 年 6 月，欧盟执委会发布了《欧洲物联网行动计划》，提出了包括监管、隐私保护、芯片、基础设施保护、标准修改、技术研发等在内的 14 项保障物联网加速发展的技术。行动方案的主要内容为：

(1) 加强物联网管理；

(2) 完善隐私和个人数据保护；

(3) 提高物联网的可信度、接受度和安全性；

(4) 评估现有物联网的有关标准并推动新标准的制定；

(5) 推进物联网方面的研发；

(6) 通过欧盟竞争力和创新框架计划(CIP)推动物联网应用；

(7) 加强对物联网发展的监测、统计和管理等。

2008 年 4 月，美国国家情报委员会(National Intelligence Council，NIC)将物联网列入“到 2025 年对美国利益具有重大影响的 6 项颠覆性民用技术”之一。2009 年 1 月，IBM 与美国信息技术与创新基金会(Information Technology and Innovation Foundation，ITIF)共同向奥巴马政府提交了题为《数字化复兴之路：创造工作、提升生产力和复兴美国》的报告，建议政府投资新一代的智能型基础设施，包括宽带网络、智能医疗和智能电网三大领域，以改善经济，增加就业，带动美国经济长期发展。同年 2 月，美国议会通过了《经济复苏和再投资行动法案》，其中包括了对上述三大领域的投资与发展计划。

我国也高度关注重视物联网的研究，2009 年 8 月，时任国务院总理温家宝关于“感知中国”的讲话把我国物联网领域的研究和应用开发推向了高潮，无锡市率先建立了“感知中国”研究中心，中国科学院、运营商、多所大学在无锡建立了物联网研究院。我国的《国家中长期科学与技术发展规划(2006—2020 年)》和“新一代宽带移动无线通信网”重大专项中均将物联网列入重点研究领域。

依靠物联网人类可以以更加精细和动态的方式管理生产和生活，达到“智慧”状态，提高资源利用率和生产力水平，改善人与自然间的关系。物联网的出现将从生活、生产、社会、经济、政治、军事、科技等方方面面影响人类生活和世界。

1.1.2 物联网体系结构

虽然物联网的定义目前没有统一的说法，但物联网的技术体系结构基本得到统一认识，分为感知层、网络层、应用层三个大层次，如图 1-1 所示。

1. 感知层

感知层是物联网的“皮肤”和“五官”，负责采集物理世界中发生的物理事件和

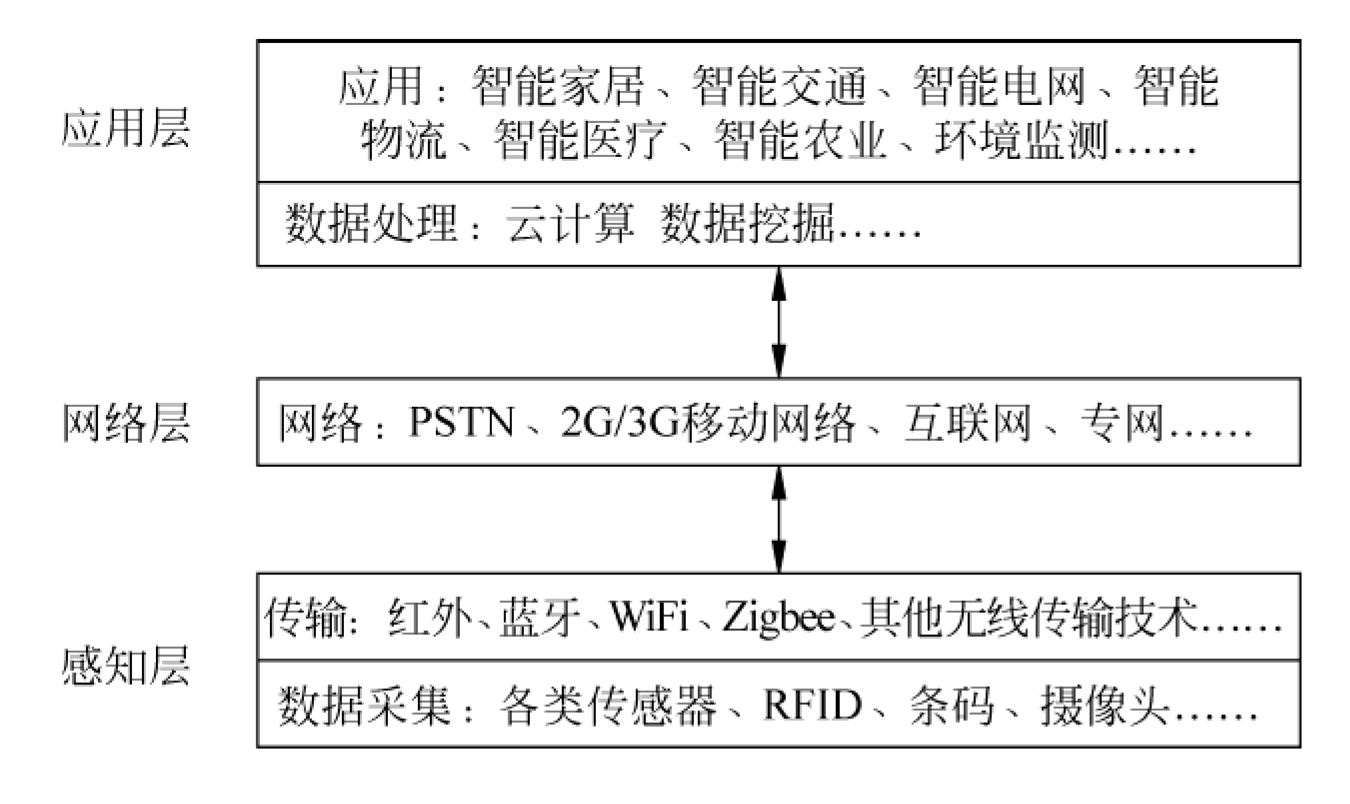


图 1-1　物联网的典型技术架构

数据，包括各类物理量、身份标识、位置信息、音频、视频数据等。针对具体感知任务，常采用协同处理的方式对多种类、多角度、多尺度的信息进行在线计算与控制，并通过接入设备将获取的信息与网络中的其他单元进行资源共享与交互。感知层主要实现物体的信息采集、捕获和识别。感知延伸层的关键技术包括传感器、RFID、GPS、自组织网络、传感器网络、短距离无线通信等。感知层的关键技术包括传感器、RFID、GPS、自组织网络、传感器网络、短距离无线通信等。感知层又分为数据采集与执行和短距离无线通信两个部分。数据采集与执行主要是运用智能传感器技术、身份识别以及其他信息采集技术，对物品进行基础信息采集，同时接收上层网络送来的控制信息，完成相应执行动作。短距离无线通信能完成小范围内的多个物品的信息集中与互通功能。

2. 网络层

网络层是物联网的神经系统，主要进行信息的传递，完成大范围的信息传输，主要通过现有的移动通信网（如 GSM 网、TD-SCDMA 网）、无线接入网（如 WiMAX）、无线局域网（WiFi）、卫星网等基础设施，把感知层感知到的信息快速、可靠、安全地传送到地球的各个地方，使物品能够进行远距离、大范围的通信，以实现在地球范围内的通信。其中，以 IPv6/IPv4 以及后 IP（Post-IP）为核心的互联网平台，将网络内的信息资源整合成一个可以互联互通的大型智能网络，为上层服务管理和大规模行业应用建立起一个高效、可靠、可信的基础设施平台。

3. 应用层

应用层完成物联网中信息处理和应用，面向各类应用，实现信息的存储、数据的挖掘、应用的决策等，涉及海量信息的智能处理、分布式计算、中间件、数据挖掘等多种技术。

由于网络层是由多种异构网络组成的，而物联网的应用是多种多样的，因此通

过中间件来承上启下。在过去的几年中,面向服务的架构(Service Oriented Architecture,SOA)是中间件实现的主流技术,通过构建在SOA基础上的服务可以以一种统一和通用的方式进行交互,实现业务的灵活扩展。在海量数据处理中,云计算是物联网智能信息分析的核心技术,它为数以亿计的各类物品的实时动态管理提供了技术基础。随着物联网应用的发展、终端数量的增长,可借助云计算处理海量信息,提升物联网信息处理能力。因此,云计算作为一种虚拟化、硬件/软件运营化的解决方案,可以为物联网提供高效的计算、存储能力。

物联网的根本还是为人服务,应用层完成物品与人的最终交互,下面两层感知并收集传输信息,交给应用层进行统一分析、决策,用于支撑跨行业、跨应用、跨系统之间的信息协同、共享、互通,提高信息的综合利用度,最大限度地为人类提供服务。

1.1.3 物联网关键技术

如图1-2所示,物联网由传感器网络、射频读写器、条码与二维码等设备以及互联网组成。依据前面的技术体系结构,可以看出,从感知层的基础传感器到应用层的海量数据整合与挖掘,以及物联网的安全、标准都存在有待解决的关键技术问题,涉及的关键技术包括信息感知与处理、短距离无线通信、广域网络、云计算、数据挖掘、安全、标准等。

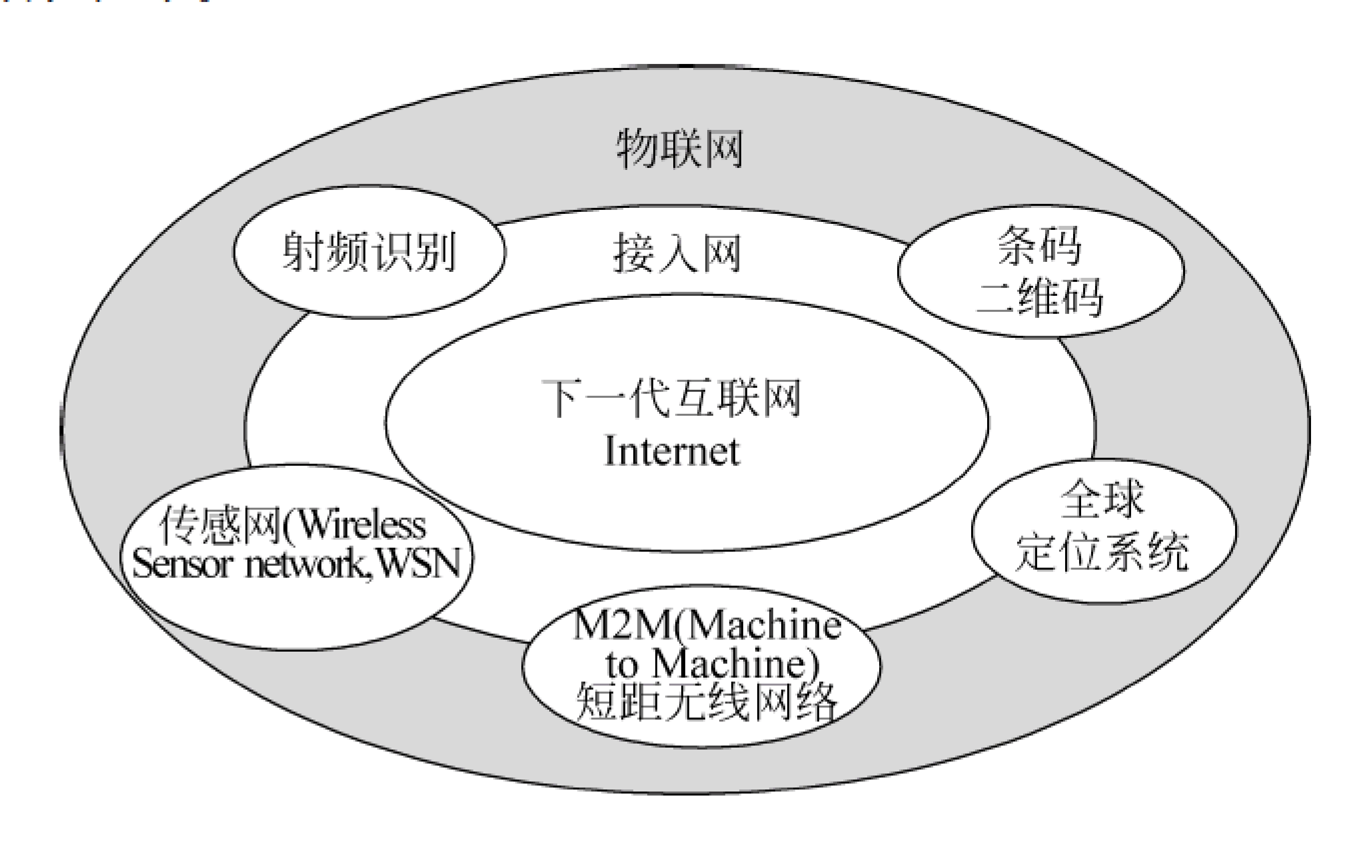


图1-2 物联网关键技术组成

1. 智能感知技术

要让物品说话,人要听懂物品的话,看懂物品的动作,感知设备是关键,是物联网实现“物物相联,人物互动”的基础。物联网的数据采集技术包括传感器技术、嵌入式系统技术、采集设备以及核心芯片。智能感知涉及3个关键问题:一是物品的种类繁多,各种各样,千差万别,物联网末端的感知设备也就种类繁多,不像电话

网、互联网的末端是针对人的，种类可以比较单一；二是物品的数量巨大，存在同一编址问题，IPv6 地址众多，但它是针对因特网设计的，对物联网终端，其复杂度、成本、功耗都是有待解决的问题；三是成本问题，物联网终端数量巨大，其成本、功耗等都有更加苛刻的要求。

2. 智能信号处理技术

采用智能信号处理技术可以对感知设备获得的各种原始数据进行处理，以获得与目标事物相关的信息。首先获得各种物理量的量测值，即原始信号；然后，通过信号提取技术筛选有用信号，并根据需要进行信号变换，在映射空间上进行信号的特征提取；最后采用模式识别、数据挖掘等技术可以将各种特征信号与某一类的物理事件相关联。

3. 短距离无线通信

短距离无线通信是感知层中非常重要的一个环节，由于感知信息的种类繁多，各类信息的传输对所需通信带宽、通信距离、无线频段、功耗、成本敏感度等都存在很大的差别，因此在短距离无线通信中，需要根据具体需求采用相应的传输网络。

4. 数据融合与挖掘

物联网中的信息种类、数量都成倍增加，其需要分析的数据量成级数规律增加，同时还需要考虑各种感知数据的融合问题，如何从海量数据中挖掘隐藏信息等问题，这都给数据计算带来了巨大挑战。

5. 安全

物联网的安全与现有信息网络的安全问题不同，它不仅包含信息的保密安全，同时还新增了信息真伪鉴别方面的安全。物联网中的传感节点通常部署在无人值守、不可控制的环境中，除了受到一般无线网络所面临的信息泄露、重放攻击、信息篡改、拒绝服务等多种威胁外，还面临传感节点容易被攻击者获取，通过物理手段获取存储在节点中的信息，从而侵入网络、控制网络的威胁。因此，在物联网安全领域，数据安全协议、密钥建立及分发机制、数据加密算法设计以及认证等技术都是关键问题所在。

6. 标准

不管哪种网络技术，标准是关键，物联网涉及的环节多，终端种类更多，其标准也更多。必须有标准，才能使各个环节的技术互通，才能融入更多的技术。在国家层面，标准更是保护国家利益和信息安全的最佳手段。

1.2 物联网编码技术

1.2.1 概念

在现实生活中,各种各样的活动或者事件都会产生这样或者那样的数据,这些数据包括人的、物质的、财务的,也包括采购的、生产的和销售的。这些数据的采集与分析对于我们的生产或者生活决策来讲是十分重要的。在计算机信息处理系统中,数据的采集是信息系统的基础,这些数据通过数据系统的分析和过滤,最终成为影响我们决策的信息。

在信息系统早期,相当多的数据处理都是通过手工录入的,这样,不仅数据量十分庞大,劳动强度大,而且数据误码率较高,缺乏实时性。为了解决这些问题,人们就研究和发展了各种各样的自动识别技术,以提高数据采集的准确性、实时性,为企业业务分析和决策提供参考依据。

自动识别技术(Automatic Identification,Auto-ID)是以计算机技术和通信技术的发展为基础的综合性科学技术,它是信息数据自动识读、自动输入计算机的重要方法和手段。

自动识别技术是数据编码、数据采集、数据标识、数据管理、数据传输的标准化手段。自动识别系统是一个以信息处理为主的技术系统,它的输入端是将被识别的信息,输出端是已识别的信息。

自动识别技术近几十年在全球范围内得到了迅猛发展,初步形成了一个包括条码技术、磁条磁卡技术、IC 卡技术、射频技术、声音识别及视觉识别等集计算机、光、磁、物理、机电、通信技术为一体的高新技术学科。而中国物联网校企联盟认为,自动识别技术可以分为光符号识别技术、语音识别技术、生物计量识别技术、IC 卡技术、条形码技术、射频识别技术(RFID)。

一般来讲,在一个信息系统中,数据的采集(识别)完成了系统的原始数据的采集工作,解决了人工数据输入的速度慢、误码率高、劳动强度大、工作简单重复性高等问题,为计算机信息处理提供了快速、准确地进行数据采集输入的有效手段。

完整的自动识别计算机管理系统包括自动识别系统(Auto Identification System,AIDS),应用程序接口(Application Interface,API)或者中间件(Middleware)和应用系统软件(Application Software)。其中,自动识别系统完成数据的采集和存储工作,应用系统软件对自动识别系统所采集的数据进行应用处理,而应用程序接口则提供自动识别系统和应用系统软件之间的通信接口,将自动识别系统采集的数据信息转换成应用软件系统可以识别和利用的信息并进行数据传递。

1.2.2　分类

自动识别技术主要包括条形码(Barcode)、智能卡(Smart Card)、射频识别(RFID)等技术,以及指纹、虹膜等生物识别技术。这些识别技术也被分为“无生命”的识别技术和“有生命”的识别技术两大类。

1. “无生命”识别技术

1）条形码识别技术

条码技术是一种最传统的自动识别技术,自 20 世纪 70 年代产生后发展至今,已经成为一种重要的信息标识和信息采集技术在世界范围内被推广应用。近年来,随着计算机应用的不断普及,条形码可以标出商品的生产国、制造厂家、商品名称、生产日期、图书分类号、邮件起止地点、类别、日期等信息,因而在商品流通、图书管理、邮电管理、银行系统等许多领域都得到了广泛的应用。

二维条码技术,即基于一维条码技术经过研究逐步兴起的一种自动识别技术。这项技术在信息容量、信息密度、中英文字符显示以及纠错等方面的功能要优于一维条码技术。在现代商业活动中应用十分广泛,如产品防伪/溯源、广告推送、网站链接、数据下载、商品交易、定位/导航、电子凭证、车辆管理、信息传递、名片交流等。如今智能手机扫一扫功能的应用使得二维码更加普遍。

2）智能卡(SmartCard)技术

智能卡实质上是一个“集成电路卡”,将具有处理能力和安全可靠、加密存储功能的集成电路芯板嵌入一个与信用卡一样大小的基片中。其最大特点是具有独立的运算和存储功能,易与计算机系统结合,在信息的采集、管理、传输、加密等方面更便于实现,被广泛应用于物流领域,例如智能货运车辆识别、物品身份追踪与验证等方面。

3）射频识别技术

射频识别技术(Radio Frequency Identification,RFID)是一种非接触式的自动识别技术。通过无线电信号识别特定目标并读写相关数据,而无须识别系统与特定目标之间建立机械或光学接触。

常用的 RFID 有低频(125～134.2kHz)、高频(13.56MHz)、超高频、微波等技术。目前 RFID 技术已经被广泛应用于物流、图书管理、门禁、零售、食品追溯、移动对象跟踪等领域。

2. “有生命”的识别技术

“有生命”的识别技术,即生物特征识别技术,是通过计算机与各种传感器和生物统计学原理等手段密切结合,利用人体固有的生理特性和行为特征,来进行个人身份鉴定的技术。

1）指纹识别技术

指纹是指人的手指末端正面皮肤上凸凹不平产生的纹线，纹线有规律的排列形成不同的纹型，纹线的起点、终点、结合点和分叉点，称为指纹的细节特征点(minutiae)。由于指纹具有终身不变性、唯一性和方便性，已经成为生物特征识别的主要技术。指纹识别是根据人体指纹的纹路、细节特征等信息对操作或被操作者进行身份鉴定的技术，已经成为目前生物特征识别技术中研究最深入、应用最广泛、发展最成熟的技术之一。

2）虹膜识别技术

人的眼睛结构由巩膜、虹膜、瞳孔三部分构成。虹膜是位于黑色瞳孔和白色巩膜之间的圆环状部分，其包含有很多相互交错的斑点、细丝、冠状、条纹、隐窝等的细节特征，这些特征决定了虹膜特征的唯一性，同时也决定了身份识别的唯一性。虹膜识别是当前应用最为方便和精确的一种技术，被认为是21世纪最具有发展前途的生物认证技术，在安防、国防、电子商务等多种领域得到了广泛应用。

3）人脸识别技术

人脸识别是基于人的脸部特征信息进行身份识别的一种生物识别技术。用摄像机或摄像头采集含有人脸的图像或视频流，并自动在图像中检测和跟踪人脸，通常也叫做人像识别、面部识别。目前，在门禁考勤管理、公安、司法、刑侦、电子商务等领域得到了广泛应用。

1.2.3 条码技术

1. 一维码

条形码是将宽度不等的多个黑条和空白，按照一定的编码规则排列，用以表达一组信息的图形标识符。常见的条形码是由反射率相差很大的黑条(简称条)和白条(简称空)排成的平行线图案。条形码在图书管理、邮政管理、商品流通、银行系统等许多领域都得到广泛的应用。

通常对于每一种物品，它的编码是唯一的，对于一维条码，一般需要通过数据库建立条码与商品信息的对应关系。

如图1-2所示，条码的主要包括以下几个部分：

通用商品条形码一般由前缀部分、制造厂商代码、商品代码和校验码组成，如图1-3所示。

(1) 国家码：是用来标识国家或地区的代码，赋码权在国际物品编码协会，如00～09代表美国、加拿大，45、49代表日本，69代表中国大陆，471代表中国台湾地区，489代表中国香港特别行政区。

(2) 生产商编码：各个国家或地区的物品编码组织来分配授权，中国由国家物

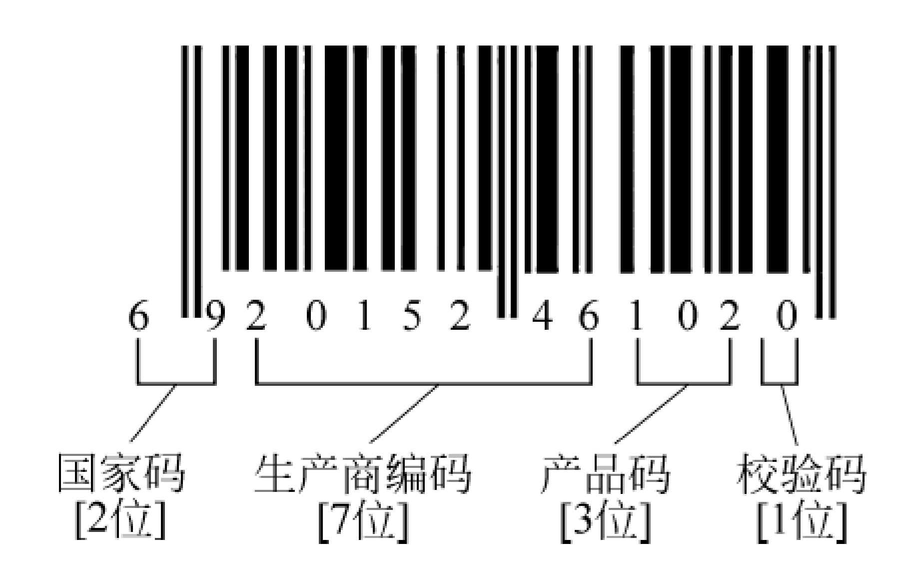


图 1-3　条码的基本结构

品编码中心赋予制造厂商代码。

(3) 产品码：是用来标识商品的代码，赋码权由产品生产企业自己行使。

(4) 校验位：最后用 1 位校验码来校验商品条形码中左起第 1～12 数字代码的正确性。

条形码校验码公式：

首先，把条形码从右往左依次编序号为"1,2,3,4……"，从序号二开始将所有偶数序号位上的数相加求和，用求出的和乘以 3，再将所有奇数序号上的数相加求和，用求出的和加上刚才偶数序号上的数，然后得出和。再用 10 减去这个和的个位数，就得出校验码。

(1) 条码编码方法有两种。

宽度调节法：指条码的条(空)宽的宽窄设置不同。用宽单元表示二进制 1，用窄单元表示二进制 0，宽窄单元比一般为 2∶1～3∶1。

模块组配法：指条码符号中每个字符的条与空分别由若干个模块组配而成，模块宽的条表示二进制 1，模块宽的空表示二进制 0。

(2) 构成条码的基本单位是模块，模块是指条码中最窄的条或空。模块的宽度通常以 mm 或 mil(千分之一英寸)为单位。构成条码的一个条或空称为一个单元，一个单元包含的模块数是由编码方式决定的。在有些码制中，如 EAN 码，所有单元由一个或多个模块组成；而另一些码制，如 39 码中，所有单元只有两种宽度，即宽单元和窄单元，其中的窄单元即为一个模块。

(3) 密度：条码的密度指单位长度的条码所表示的字符个数。对于一种码制而言，密度主要由模块的尺寸决定，模块尺寸越小，密度越大，所以密度值通常以模块尺寸的值来表示(如 5mil)。通常 7.5mil 以下的条码称为高密度条码，15mil 以上的条码称为低密度条码，条码密度越高，要求条码识读设备的性能(如分辨率)也越高。高密度的条码通常用于标识小的物体(如精密电子元件等)，低密度条码一般应用于远距离阅读的场合(如仓储管理、图书、商品销售等)。

(4) 宽窄比。

对于只有两种宽度单元的码制,宽单元与窄单元的比值称为宽窄比,一般为2～3(常用的有2∶1和3∶1)。宽窄比较大时,阅读设备更容易分辨宽单元和窄单元,因此比较容易阅读。

(5) 对比度(Print Contrast Signal,PCS):条码符号的光学指标,PSC值越大则条码的光学特性越好。

PCS=(RL－RD)/RL×100%(RL:条的反射率,RD:空的反射率)

(6) 条码字符集。

条码字符集指某种条码所含全部条码字符的集合。条码字符中字符总数不能大于该种码制的编码容量。

(7) 条码的连续性与非连续性。

连续性是指每个条码字符之间不存在间隔;相反,非连续性是指每个条码字符之间存在间隔。连续性条码密度相对较高,非连续性条码密度较低。

(8) 定长条码与非定长条码。

定长条码是指仅能表示固定字符个数的条码;非定长条码是指能表示可变字符格式的条码。由于限制了字符个数,定长条码译码误读率相对较低;非定长条码具有灵活、方便等优点,但译码误读率较高。

(9) 条码双向可读性。

条码的双向可读性是指从条码的左、右两侧开始扫描都可被识读的特性。对于双向可读的条码,识读过程译码器需要判别扫描方向。

(10) 条码的码制。

条码的码制是指条码符号的类型,不同类型的条码符号,条、空图案对数据的编码方法各有不同。每种码制都具有固定的编码容量和所规定的条码字符集。

常用的一维码码制有EAN码、UPC码、交叉25码、39码、128码以及库德巴码(Codabar码)等,不同的码制有其各自的应用领域。

- EAN码:全球推广应用的商品条码,是定长的纯数字条码,有EAN-13、EAN-8码。
- UCC/EAN-128码:一种连续型非定长条码,是唯一能够表示应用标识的条码符号,能够更多地标识贸易单元中需表示的信息,如产品批号、规格、生产日期、有效期等。
- UPC码:美国制定的在北美地区应用的定长、纯数字型码,在技术上与EAN码完全一样,有5种版本,常用的商品条码版本为UPC-A码和UPC-E码。
- 交插25码:高密度的物流码,第一个数字由条开始,第二个数字由空组成。应用于商品批发、仓库、机场、生产/包装识别。

- 39 码：一种可表示数字、字母等信息的条码。主要用于工业、图书及票证等方面的自动化管理，目前使用极为广泛。
- 库德巴码(Codabar 码)：也可表示数字和字母信息的条码，主要用于医疗卫生、图书情报、物资等领域。
- 25 码：主要应用于包装、运输以及国际航空系统的机票顺序编号等。
- ISBN：用于图书管理。

(11) 条码符号的组成。

如图 1-4 所示，一个完整的条码的组成次序依次为静区(前)、起始符、数据符、中间分割符(主要用于 EAN 码)、终止符、静区(后)。



图 1-4　条码符号的组成结构

- 静区：指条码左右两端外侧与空的反射率相同的限定区域，它能使阅读器进入准备阅读的状态，当两个条码相距较近时，静区有助于对它们加以区分，静区的宽度通常应不小于 6mm(或 10 倍模块宽度)。
- 起始/终止符：指位于条码开始和结束的若干条与空，标志条码的开始和结束，同时提供了码制识别信息和阅读方向的信息。
- 数据符：位于条码中间的条、空结构，它包含条码所表达的特定信息。

2. 二维码

由于条码技术具有输入速度快、准确度高、成本低、可靠性强等优点，因此在各行业得到了广泛应用。但随着应用领域的不断扩展，传统的一维条码渐渐表现出了它的局限：首先，使用一维条码，需要通过数据库来提取条码所表达的信息；其次，一维条码表达的只能为字母和数字，而不能表达汉字和图像，在一些需要应用汉字的场合，一维条码便不能很好地满足要求；另外，在某些场合下，大信息容量的一维条码通常受到标签尺寸的限制，也给产品的包装和印刷带来了不便。

二维条码的诞生解决了一维条码不能解决的一些问题，它能够在横向和纵向两个方位同时表达信息，不仅能在很小的面积内表达大量的信息，而且能够表达汉字和存储图像。二维条码的出现拓展了条码的应用领域，因此被许多不同的行业所采用。

1）概述

二维条码(2-dimensional barcode)是用某种特定的几何图形按一定规律在平面(二维方向上)分布的黑白相间的图形记录数据符号信息的,在代码编制上利用构成计算机内部逻辑基础的“0”、“1”比特流的概念,使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图像输入设备或光电扫描设备自动识读以实现信息自动处理。

二维条码具有如下特点：

(1）高密度编码,信息容量大,可容纳多达1850个大写字母或2710个数字或1108个字节,或500多个汉字,比普通条码信息容量约高几十倍。

(2）编码范围广,可以对图片、声音、文字、签字、指纹等可以数字化的信息进行编码,用条码表示出来；可以表示多种语言文字,可表示图像数据。

(3）容错能力强,具有纠错功能：对局部损坏的二维条码,依然可以正确识读。

(4）译码可靠性高,误码率不超过千万分之一。

(5）可引入加密措施,保密性、防伪性好。

(6）成本低,易制作,持久耐用。

(7）条码符号形状、尺寸大小比例可变。

(8）二维条码可以使用激光或CCD阅读器识读。

2）二维条码分类

二维条码可以分为堆叠式/行排式二维条码和矩阵式二维条码。

(1）堆叠式/行排式二维条码(又称堆积式或层排式),其编码原理是建立在一维条码基础之上,按需要堆积成二行或多行。在编码设计、校验原理、识读方式等方面继承了一维条码的一些特点,识读设备与条码印刷与一维条码技术兼容。

(2）矩阵式二维码(又称棋盘式二维条码)是在一个矩形空间通过黑、白像素在矩阵中的不同分布进行编码。在矩阵相应元素位置上,用点(方点、圆点或其他形状)的出现表示二进制“1”,点的不出现表示二进制的“0”,点的排列组合确定了矩阵式二维条码所代表的意义。矩阵式二维条码是建立在计算机图像处理技术、组合编码原理等基础上的一种新型图形符号自动识读处理码制。具有代表性的矩阵式二维条码有QR Code、Code One、Maxi Code、Data Matrix等。在目前几十种二维条码中,常用的码制有QR Code、PDF 417二维条码、Datamatrix二维条码、Maxicode二维条码、Code 49、Code 16K、Code one等。除此之外,还有Vericode条码、CP条码、Codablock F条码、田字码、Ultracode条码,Aztec条码。

3）常见的二维条码

(1）Code 49码：是一种多层、连续型、可变长度的条码符号。它由2～8层组成,可表示全部的128个ASCII字符,每层有18个条和17个空,层与层之间由一

个层分隔条分开，如图 1-5 所示。

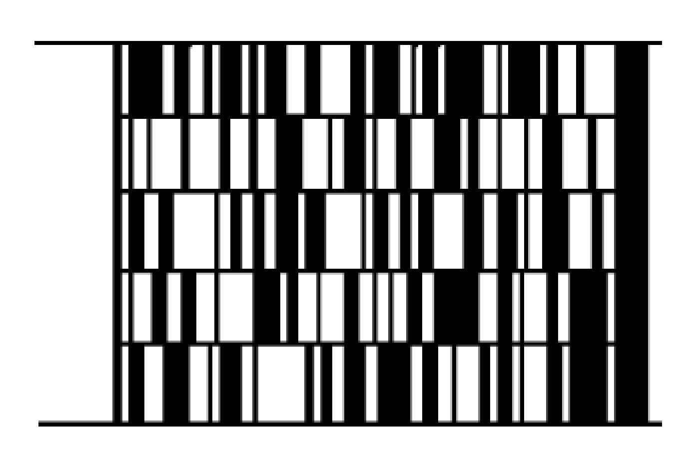


图 1-5　Code 49 码

（2）Code 16K 码：是一种多层、连续型、可变长度的条码符号。它可以表示 ASCII 所有的 128 个字符及扩展 ASCII 字符，如图 1-6 所示。

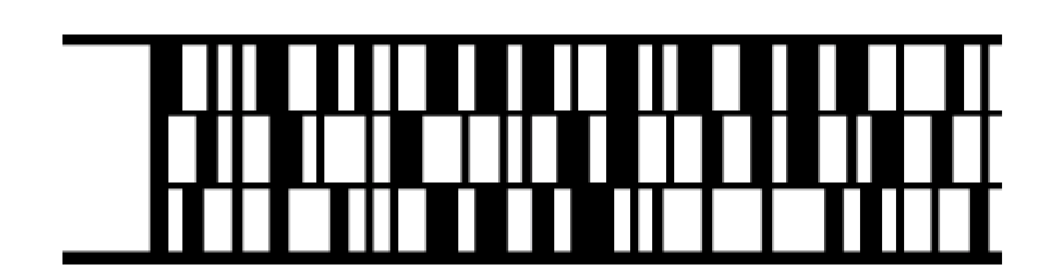


图 1-6　Code 16K 码

（3）PDF（Portable Data File，便捷数据文件）417 码：是一种多层、可变长度、具有高容量和纠错能力的二维码。它可以表示超过 1100 个字节、1800 个 ASCII 字符或 2700 个数字的数据，可通过线性或二维成像设备识读，其形式如图 1-7 所示。



图 1-7　PDF 417 码

（4）Code one 码：是一种由成像设备识别的矩阵式二维码，条码符号中包含可由快速线性探测器识别的图案，共有 10 个版本及 14 种尺寸，其形式如图 1-8 所示。

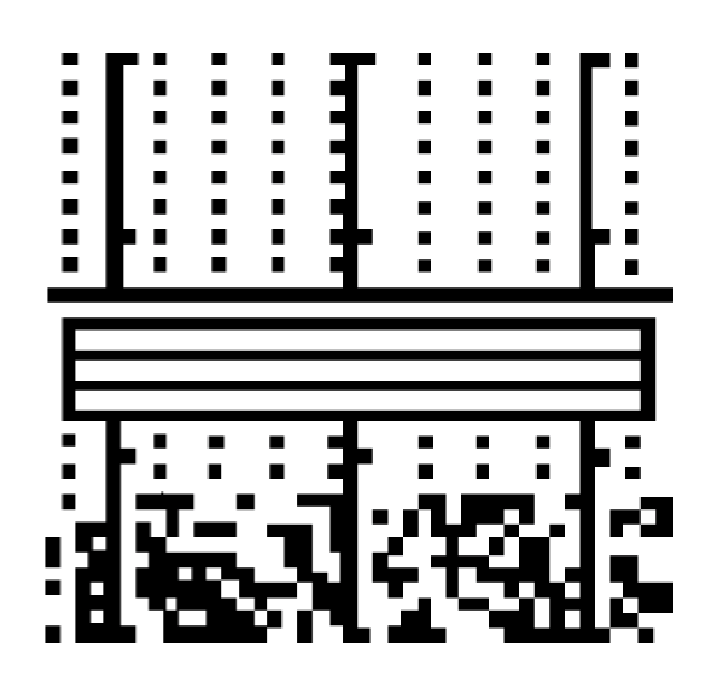


图 1-8　Code One 码

（5）Data Matrix 码：是矩阵式二维码，有两种类型，即 ECC000-140 和 ECC200。可表示全部 ASCII 字符及扩展 ASCII 字符，最大数据容量为 2335 个文本字符、2116 个数字或 1556 个字节，其形式如图 1-9 所示。

（6）Maxicode 码：是一种固定长度的矩阵二维码，由紧密相连的多行六边形模块和位于符号中央位置的定位图形组成，共有 7 种模式，可表示全部 ASCII 字符和扩展 ASCII 字符，最大数据容量为 93 个文本字符或 138 个数字，其形式如图 1-10 所示。

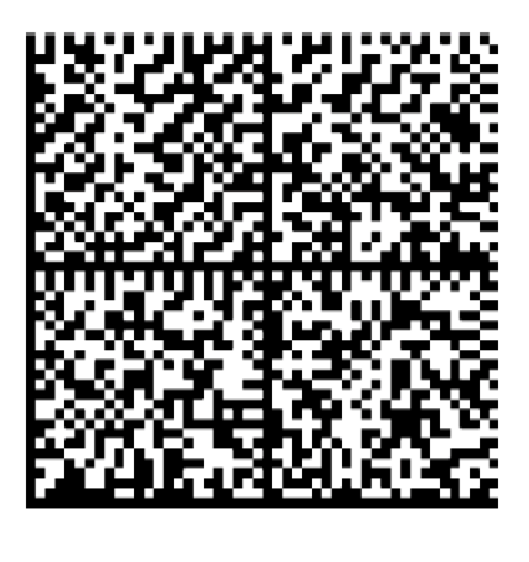


图 1-9 Data Matrix 码

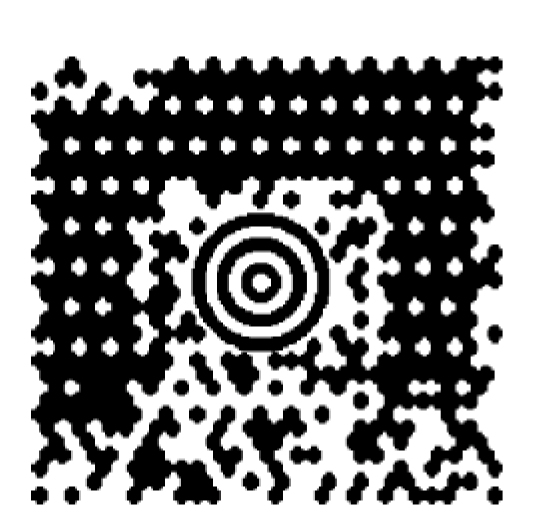


图 1-10 Maxicode 码

（7）QR 码：是日本电装公司在 1994 年向世界公布的快速响应矩阵码的简称，其可容纳大量信息，可表示数字数据 7089 个字符，可对英文、数字、汉字进行编码。即使损坏或污损也可读取，具有识读速度快、数据密度大、占用空间小的优势，其形式如图 1-11 所示。

（8）汉信码：是中国物品编码中心研制的具有自主知识产权的矩阵二维码，其具有汉字编码能力强、抗污损、抗畸变识读能力、识读速度快、信息密度高、纠错能力强、图形美观等优点，是一种适合在我国广泛应用的二维码，其形式如图 1-12 所示。



图 1-11 QR 码

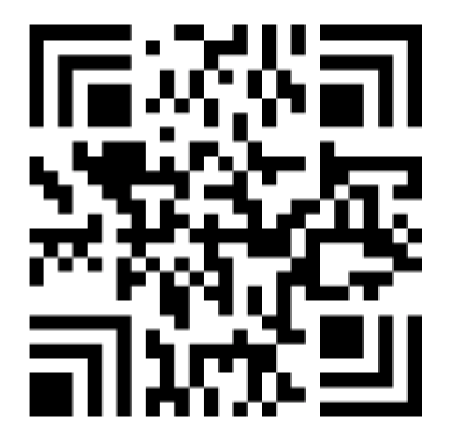


图 1-12 汉信码

4）二维码的应用

目前，二维码在移动应用中得到了广泛的应用，其与手机结合的基本原理是：信息存储在二维码中，手机通过摄像头或者彩信的方式获取。然后，根据不同的用途，或者由外置的条码阅读设备读取手机屏幕上显示的二维码，或者通过手机本身

的解码器获得相关的信息，最后再根据不同目的和程序完成后续应用。

3. 条码识别

条码的识别包括两个过程，即扫描和译码。由于物体的颜色是由其反射光的类型来决定的，因此条码的识别主要依据“白色物体能反射各种波长的可见光，黑色物体则吸收各种波长的可见光”，当条形码扫描器光源发出的光在条形码上反射后，反射光通过扫描器内部的光电转换器，将反射光线的明暗转换成数字信号。电信号输出到条码扫描器的放大电路增强信号之后，再送到整形电路将模拟信号转换成数字信号。白条、黑条的宽度不同，相应的电信号持续时间长短也不同。然后译码器通过测量脉冲数字电信号0、1的数目来判别条和空的数目。通过测量0、1信号持续的时间来判别条和空的宽度。此时所得到的数据再根据对应的编码规则（例如，EAN-8码），转换成相应的数字、字符信息。最后，由计算机系统进行处理，物品的详细信息便被识别出来了。

不论是采取何种规则印制的条形码，都由静区、起始字符、数据字符与终止字符组成。有些条码在数据字符与终止字符之间还有校验字符。

（1）静区：静区也叫空白区，分为左空白区和右空白区，左空白区用于提示扫描设备准备开始扫描，右空白区是保证扫描设备正确识别条码的结束标记。为了防止左右空白区（静区）在印刷排版时被无意中占用，可在空白区加印一个符号作为静区标记（左侧没有数字时印“＜”号，右侧没有数字时加印“＞”号）。主要作用就是防止静区宽度不足。

（2）起始字符：第一位字符，具有特殊结构，当扫描器读取到该字符时，便开始正式读取代码了。

（3）数据字符：条形码的主要内容。

（4）校验字符：检验读取到的数据是否正确。不同的编码规则采用不同的校验规则。

（5）终止字符：最后一位字符，用于提示代码扫描完毕，同时还起到进行校验计算的作用。

为了方便双向扫描，起止字符具有不对称结构。因此扫描器扫描时可以自动对条码信息重新排列。条码扫描器有光笔、CCD、激光和影像四种。

（1）光笔：是最原始的扫描方式，需要手动移动光笔，并且还要与条形码接触。

（2）CCD：以CCD作为光电转换器，LED作为发光光源的扫描器。在一定范围内，可以实现自动扫描。并且可以阅读各种材料、不平表面上的条码，成本也较为低廉，但扫描距离较短。

（3）激光：是一种光学距离传感器，以激光作为发光源的扫描器，扫描方式有单线扫描、光栅式扫描和全角度扫描三种方式。

（4）影像：以光源拍照利用自带硬解码板解码，通常影像扫描可以同时扫描一维及二维条码，如 Honeywell 引擎。

1.2.4 EPC 编码

EPC(Electronic Product Code)即电子产品编码，是一种编码系统。EPC 是由美国麻省理工学院成立的 Auto-ID 中心于 1999 年在美国统一代码委员会(Uniform Code Council，UCC)支持下，将 RFID 技术与因特网结合，提出的一种电子产品代码。

由欧洲物品编码协会(European Article Numbering Association，EAN)发展而来的国际物品编码协会(EAN International)，与美国统一代码委员会 UCC 将全球统一标识编码体系植入 EPC 概念中。2003 年 11 月 1 日，国际物品编码协会正式接管了 EPC 在全球的推广应用工作，成立了一个非营利组织 EPCglobal，负责全球 EPC 的管理和实施工作，在全球范围内对各个行业建立和维护 EPCglobal 网络，保证供应链各环节信息的自动、实时识别，采用全球统一标准。EPCglobal 网络是实现自动、即时识别和供应链信息共享的网络平台。通过整合现有信息系统和技术，EPCglobal 网络将提供对全球供应链上贸易单元即时、准确、自动的识别和跟踪。

目前，EPCglobal 属于国际物品编码协会(Global Standards 1，GS1)。GS1(即原来的 EAN International)致力于建立“全球统一标识系统和通用商务标准——EAN・UCC 系统”，通过向供应链参与方及相关用户提供增值服务，优化全球供应链的管理效率。目前，GS1 已有遍及世界 100 多个国家和地区的 100 余个系统成员，负责组织实施当地的 EAN/UCC 系统推广应用工作。

EPCglobal 的系统成员分为两类：终端成员和系统服务商。终端成员包括制造商、零售商、批发商、运输企业和政府组织。通常来说，终端成员是指在供应链中有物流活动的组织，而系统服务商是指那些给终端用户提供供应链物流服务的组织机构，包括软件和硬件厂商、系统集成商和培训机构等。目前全世界已有 100 多个国家和地区的超过 100 万家企业，使用该系统对物品进行标识和供应链管理。

1. EPCglobal 的组织架构及标准体系

1）EPCglobal 的组织架构

EPCglobal 的组织架构如图 1-13 所示。EPCglobal 管理委员会为 EPCglobal 的决策机构，由来自 UCC、EAN、MIT、终端用户和系统集成商的代表组成。下设四个委员会、Auto-ID 实验室和多个商务与技术行动组，负责规划商务和技术愿景，以促进标准发展进程。每个行动组又下设若干个工作组，作为行动组执行其事务的具体组织。

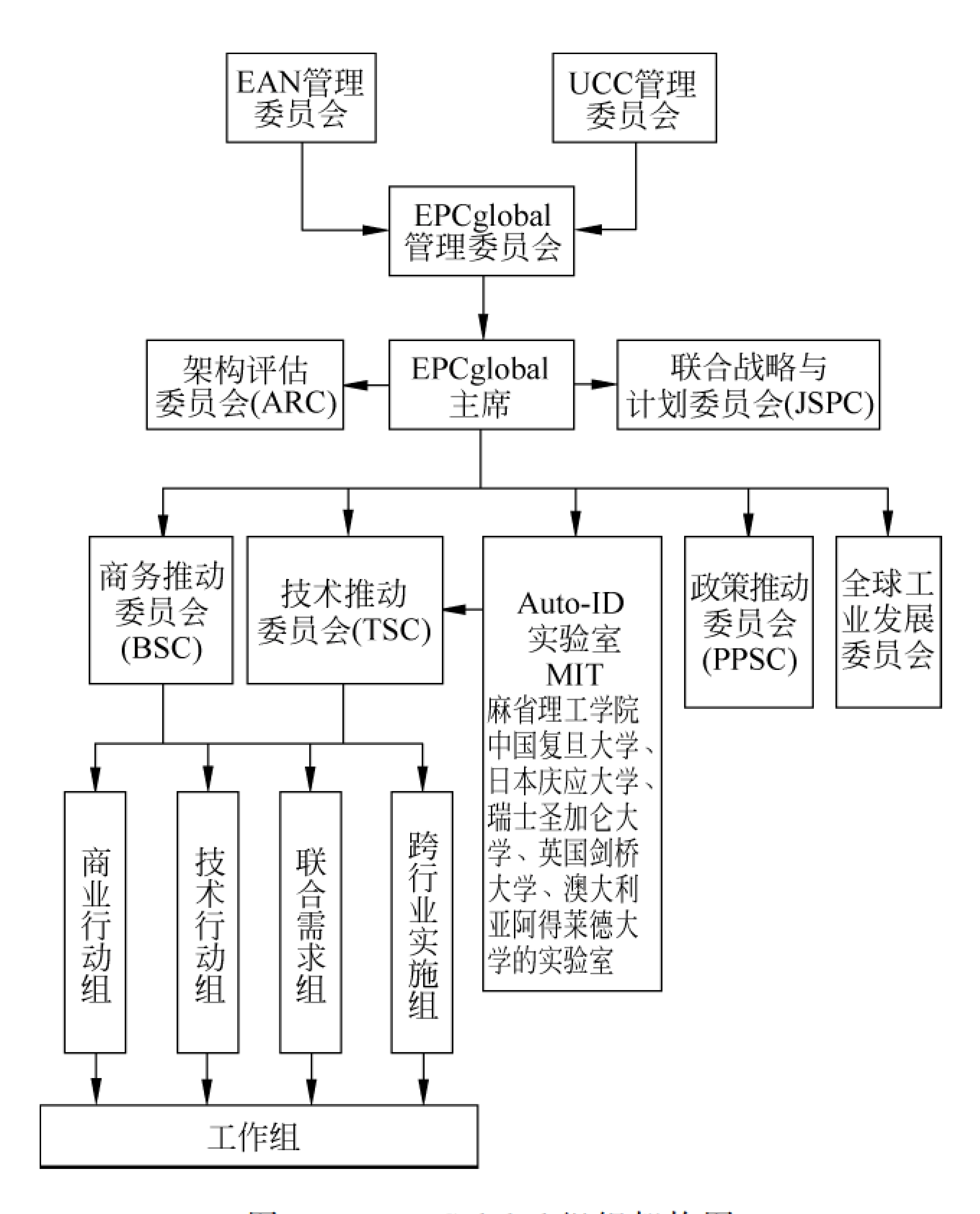


图 1-13 EPCglobal 组织架构图

其中，架构评估委员会(Architectural Review Committee，ARC)为 EPCglobal 管理委员会提供技术支持，向 EPCglobal 主席做出报告，从整个 EPCglobal 的相关构架来评价和推荐重要的需求；联合战略与计划委员会(Joint Strategy and Planning Committee，JSPC)为 EPCglobal 用户制定跨行业的应用愿景，确保 EPCglobal 的战略和工作计划可以反映每个行业最新的应用需求；商务推动委员会(Business Steering Committee，BSC)针对终端用户的需求以及实施行动来指导所有商务行动组和工作组；国家政策推动委员会(Public Policy Steering Committee，PPSC)负责对所有行动组和工作组的国家政策发布(如安全隐私等)进行筹划和指导；技术推动委员会(Technology Steering Committee，TSC)则负责对所有工作组所从事的软件、硬件和技术活动进行筹划和指导；Auto-ID 实验室是由 Auto-ID 中心发展而成的，总部设在美国麻省理工学院，与其他五所大学(分别是英国剑桥大学、澳大利亚阿德莱德大学、日本庆应大学、中国复旦大学和瑞士圣加仑大学)合作研究和开发 EPCglobal 网络及其应用。

四类行动组包括工业行动组(Industry Action Group，IAG)，由加入

EPCglobal 网络的公司代表组成，目标是完善多个工业间的商业需求，并通过实际案例推动成员的供应链效率；技术行动组(Technical Action Group，TAG)又分为制定 RFID 硬件设备间接口标准的硬件行动组(Hardware Action Group，HAG)和定义 EPCglobal 网络的软件功能及接口标准的软件行动组(Software Action Group，SAG)；联合需求组(Joint Requirements Group，JRG)的目标是寻找和利用跨行业的 EPC 应用需求间的共同点，联合提出更加统一和详细的商业和用户需求，并提交给硬件行动组和软件行动组；跨行业实施组(Cross Industry Adoption & Implementation Group，AIG)的目标是驱动全球和地区性的 EPC 应用，讨论和排除战略计划实施中出现的障碍，在成员间推广和巩固最佳经验。在四类行动组下，目前共有 40 多个活跃的工作组组织和开展 EPCglobal 活动。

2）EPC 的 RFID 标准体系

EPCglobal 的 RFID 标准体系框架包含硬件、软件、数据标准，以及由 EPCglobal 运营的网络共享服务标准等多个方面的内容。EPCglobal 标准框架如图 1-14 所示，包括数据识别、数据获取和数据交换三个层次。其中，数据识别层的标准包括 RFID 标签数据标准和协议标准，目的是确保供应链上的不同企业间数据格式和说明的统一性；数据获取层的标准包括读写器协议标准、读写器管理标准、读写器组网和初始化标准，以及中间件标准等，定义了收集和记录 EPC 数据的主要基础设施组件，并允许最终用户使用具有互操作性的设备建立 RFID 应用；数据交换层的标准包括 EPC 信息服务标准(EPC Information Services，EPCIS)、核心业务词汇标准(Core Business Vocabulary，CBV)、对象名解析服务标准(Object Name Service，ONS)、发现服务标准(Discovery Services)、安全认证标准(Certificate Profile)，以及谱系标准(Pedigree)等，目的是为最终用户提供可以共享的 EPC 数据，并实现 EPC 网络服务的接入。

上述三个层次中的标准又分为三类：数据标准，包括安全认证标准、谱系标准、RFID 标签数据标准和 RFID 标签数据格式；接口标准，包括对象解析服务、EPC 信息服务、中间件、组网和初始化、读写器管理、底层读写器协议、读写器协议和 RFID 标签协议 UHF Class 1 Gen2；未完成的标准，包括发现服务、核心业务词汇和 RFID 标签协议 HF Class 1 Gen2。

2. EPC 系统组成

EPC 系统由 EPC 编码体系、射频识别系统及信息网络系统三部分组成。其中，信息网络系统包括 EPC 中间件、对象名称解析服务(ONS)、实体标记语言(Physical Markup Language，PML)和 EPC 信息服务(EPCIS)。表 1-1 列出了 EPC 系统的构成。

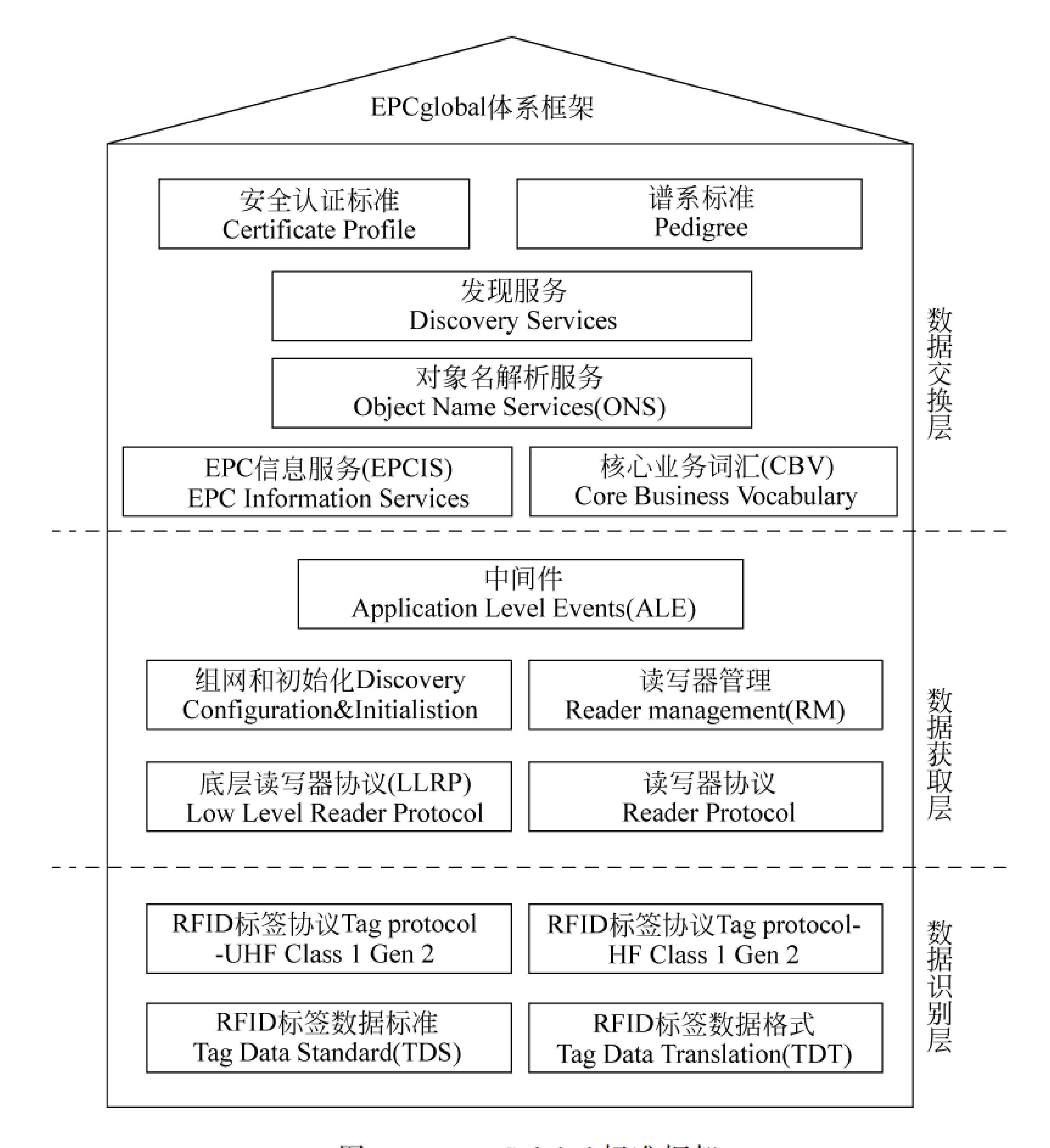


图 1-14 EPCglobal 标准框架

表 1-1 EPC 系统的构成

系统构成	名称	备注
EPC 编码体系	EPC 编码标准	识别物品的特定编码
射频识别系统	EPC 标签	贴在物品上
	读写器	识读 EPC 标签
信息网络系统	EPC 中间件	EPC 软件支持系统
	对象名称解析服务(ONS)	
	实体标记语言(PML)	
	EPC 信息服务(EPCIS)	

3. EPC 编码体系

全球电子产品编码(EPC)体系是新一代的与 EAN. UCC 编码兼容的编码标准,是全球统一标识系统的延伸和拓展,是该系统的重要组成部分,也是 EPC 系统的核心与关键。EPC 编码体系与现行的全球贸易项目代码(Global Trade Item Number,GTIN)相结合,由现行的条码标准逐渐过渡到 EPC 标准,或者在未来的供应链中 EPC 和 EAN. UCC 系统共存。

EPC 编码可以为各类物体、货物、资产、位置等需要追踪的实体分配一个唯一编码。电子产品编码贯穿于供应链数据流上的所有节点,并作为构成 EPCglobal 网络中所有标准和接口的基本元素。与 IP 地址在互联网中用来标识位置相似,通过使用计算机网络,EPC 编码可用来标识和访问单个物品。射频标签中只存储 EPC 码的信息,这种做法已经得到 UCC 和 EAN 两个机构的大力支持。

EPC 编码是由版本号、域名管理、对象种类、序列号组成的一组数字。其中,版本号标识 EPC 的版本号,它使得以后的 EPC 可有不同的长度或类型;域名管理是描述与此 EPC 相关的生产厂商的信息;对象种类类型是描述生产厂商生产的产品型号;序列号是唯一标识物品的编号。

目前,已有的 EPC 编码体系有 EPC-64、EPC-96 和 EPC-256,如表 1-2 所示。出于成本因素的考虑,参与 EPC 测试所使用的编码标准是 64 位数据结构,实际应用则采用 96 位编码结构,未来将采用 256 位编码结构。

表 1-2 EPC 的几种编码结构

名称	类型	版本号	域名管理者	对象分类代码	序列号
EPC-64	1	2 位	21 位	17 位	24 位
	2	2 位	15 位	13 位	34 位
	3	2 位	26 位	13 位	23 位
EPC-96	1	8 位	28 位	24 位	36 位
EPC-254	1	8 位	32 位	56 位	160 位
	2	8 位	64 位	56 位	128 位
	3	8 位	128 位	56 位	64 位

EPC-64 1 型编码提供了 2 位版本号编码,提供了 21 位域名管理编码,提供了 17 位对象分类编码与 24 位序列号。其中,域名管理者字段允许 2 000 000 个生产厂商使用该类型编码;对象分类字段可以容纳 131 072 个产品种类,因此绝大多数生产厂商的需求能够得到满足;序列号字段可以标识 16 000 000 个独立的产品个体。因此,普通生产商适合使用 EPC-64 1 型编码。

EPC 编码的分配由 EAN. UCC 组织管理。在中国,EAN. UCC 中 GTIN 编码由中国物品编码中心负责分配和管理,该中心同时也是 EPCglobal 及 GS1 在中国

的代表机构。

在电子产品编码分配机构向 EPC 管理者授权时，首先为 EPC 管理者分配一个唯一代码，即 EPC 管理者代码。一个 EPC 用户可以同时拥有多个 EPC 管理者代码，以此管理和维护多个 EPC 编码段。在电子产品编码的定义中，EPC 管理者代码作为独立的一部分，这样就可以通过电子产品编码直接识别 EPC 管理者的信息，以保证系统的可扩展性。举例来说，一个 ONS 查询可以从概念上理解为在一个大表中查询某个电子产品编码所映射到的 EPCIS 服务地址，但是假如有了 EPC 管理者代码，就可以由 EPC 管理者负责维护 ONS 服务器中所分配编码段的小表，这样就可以提高执行 ONS 查询的效率。

4. EPC 射频识别系统

物品编码可存储在物品的电子标签中，由读写器对电子标签进行读写，电子标签与读写器构成一个识别系统。射频识别系统实现了物品数据采集的完全自动化，是物联网的重要环节。

射频电子标签是产品电子代码（EPC）的物理载体，主要由天线和芯片组成。EPC 标签中存储的唯一信息是 96 位或者 64 位产品电子代码。为了降低成本，EPC 标签通常是被动式射频标签。它可以附着于可跟踪的物品上，可全球流通并对其进行识别和读写。它采用 RFID 技术，对每个实体对象，包括集装箱、零售商品等提供唯一性标识。同条形码技术相比，EPC 标签具有更多的优点，如信息容量更大、应用更灵活、抗干扰和抗环境污染等。

根据版本号和基本功能的不同，EPC 标签有代（Gen）和类（Class）的概念，Gen 是指 EPC 标签规范的版本号，Class 描述的是 EPC 标签的基本功能。EPCglobal 第一代标准，即 EPC Gen1 标准是 EPC 射频识别技术的基础。EPC 标签可以分为 Class 0、Class 1、Class 2、Class 3 和 Class 4 共五个类别。读写器用于识别 EPC 标签，并与信息系统相连实现数据的交换。读写器读取 EPC 标签信息的方式有多种方式，近距离读取被动标签最常用的方法是电感耦合方式。当 EPC 标签靠近读写器时，读写器的天线与标签的天线之间会形成一个磁场。标签利用这个磁场发送电磁波给读写器，返回的电磁波被转换为数据信息，也就是标签中包含的 EPC 代码。

5. EPC 信息网络

1）EPC 系统的工作流程

EPC 系统的信息网络系统是在全球互联网的基础上，通过 EPC 中间件、对象名称解析服务（ONS）和 EPC 信息服务（EPC Information Services，EPCIS）来实现全球“实物互联”。

如图 1-15 所示，在 EPC 系统中读写器读取电子标签中的 EPC 代码，然后将

EPC代码传给EPC中间件处理。在电子标签上只有一个EPC代码，如果要获取与该EPC代码匹配的物品信息，就需要ONS来提供一种自动化的网络数据库服务，EPC中间件将EPC代码传给ONS，ONS指示EPC中间件到一个保存着产品信息的服务器（EPCIS）中查找，结果将以PML的格式返回给EPC中间件，并被传给用户应用。

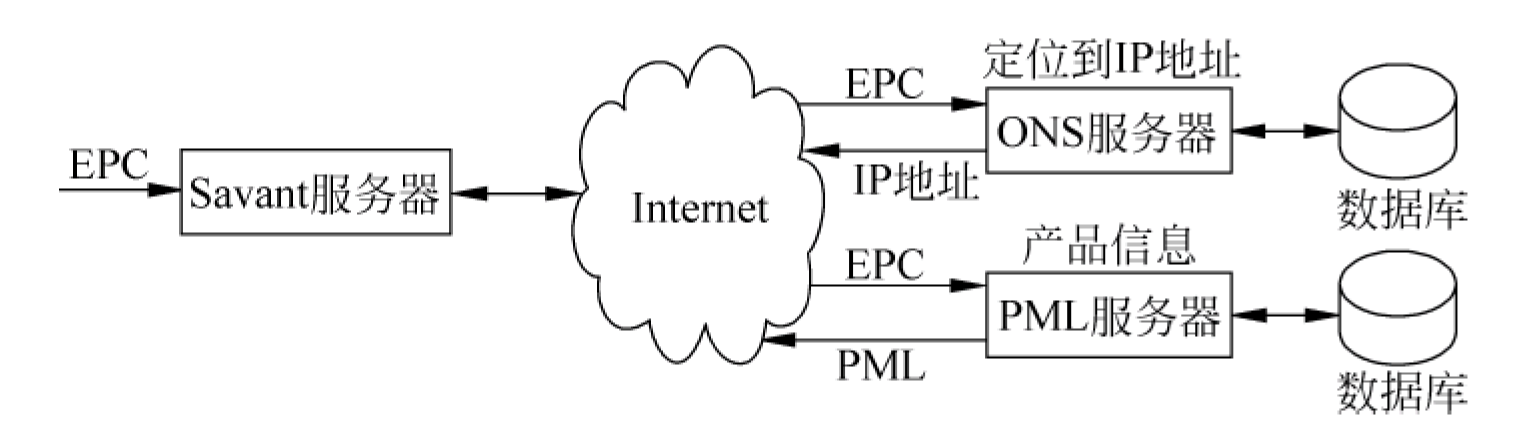


图 1-15　EPC系统工作流程图

2）EPC中间件

EPC中间件（Middleware）是物联网的神经系统，是具有一系列特定属性的“程序模块”或“服务”，并被用户集成以满足他们的特定需求，EPC中间件以前被称为Savant。EPC中间件是加工和处理来自读写器的所有信息和事件流的软件，是连接读写器和企业应用系统的纽带，主要任务是对读写器读出的EPC编码进行传送和管理。它利用了一个分布式的结构，层次化的进行组织和管理数据流。每个层次上的Savant系统将收集、存储和处理信息，并与其他的Savant系统进行交流。

美国麻省理工学院自动识别中心（Auto-ID Center）于2003年9月发布“Auto_ID Savant Specification 1.0”，作为中间件技术规范架构。该规范针对应用EPC（Electronic Products Code）的议题，确定了EPC网络技术构架。该技术架构包括Savant、实体标记语言（PML）、对象名称解析服务（ONS）。其中，Savant系统是连接读写器与企业应用系统的纽带，在将数据送往企业应用程序之前，它要对标签数据进行过滤、汇总和统计，压缩数据容量，因此，Savant系统相当于EPC网络的神经系统。图1-16描述了Savant组件与其他应用程序的通信过程。

3）对象名称解析服务（ONS）

对象名称解析服务（ONS）是一个自动的网络服务系统，类似于域名解析服务（Domain Name System，DNS），ONS给EPC中间件指明了存储产品相关信息的服务器。ONS服务是联系EPC中间件和EPC信息服务的网络枢纽，并且ONS设计与架构都以因特网域名解析服务DNS为基础，因此，可以使整个EPC网络以因特网为依托，迅速架构并顺利延伸到世界各地。

如图1-17所示，在一个局域网内的标签读写器在物理空间上分布在多个地方，用于识读不同环境的EPC标签，读写器再将读到的EPC编码信息通过局域网

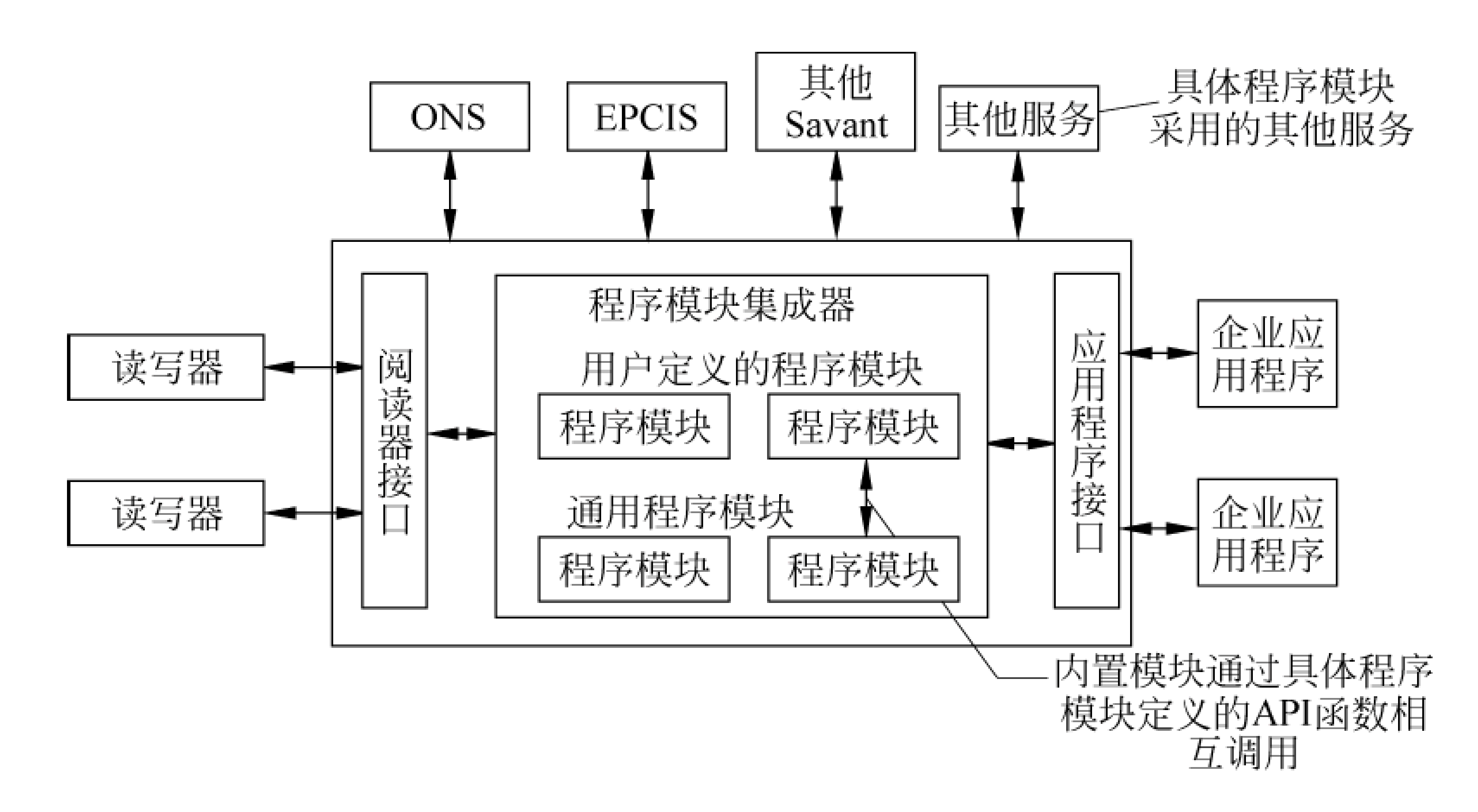


图 1-16 Savant 组件及与其他应用程序通信

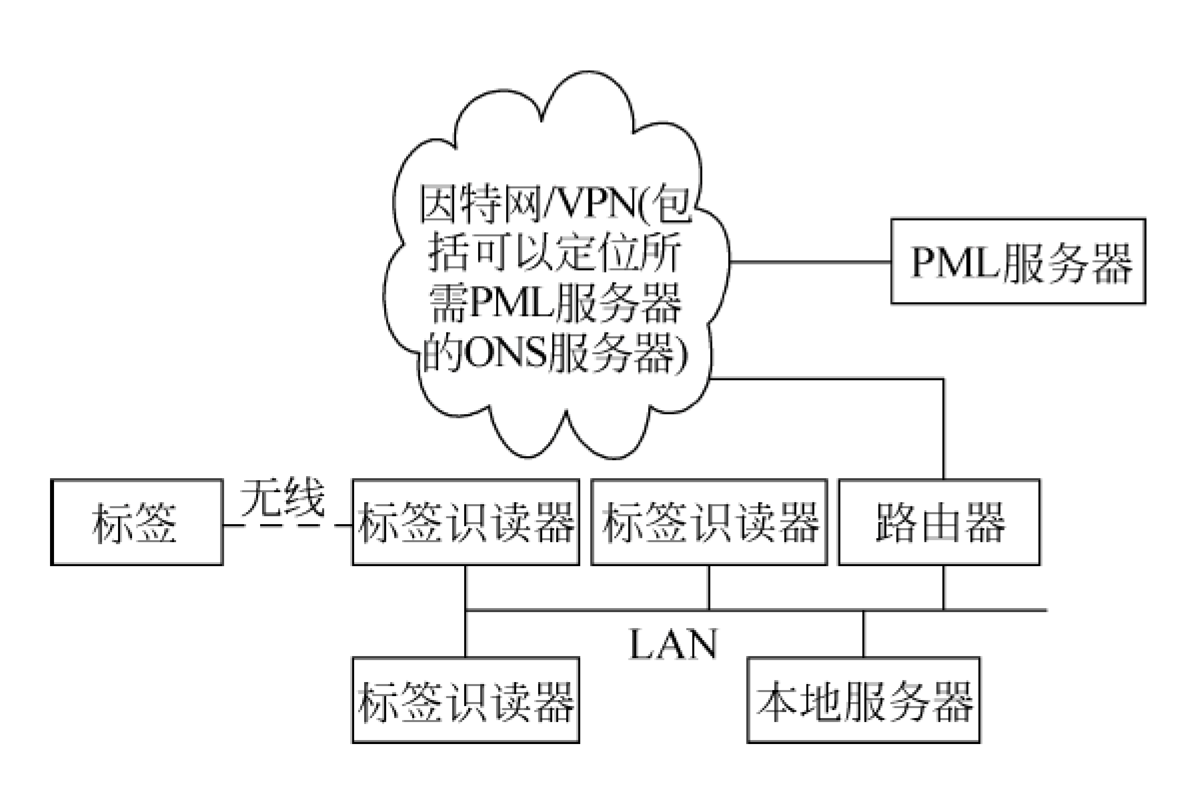


图 1-17 EPC 技术网络图

上传到本地服务器，由服务器中的 Savant 软件对这些数据进行集中处理；然后，由本地服务器通过查找本地 ONS 服务或通过路由器到达远程 ONS 服务器查找所需 EPC 编码对应的 PML 服务器地址，本地服务器就可以与找到的 PML 服务器建立连接，获取产品信息。

ONS 可在全网进行信息查找，这一点与 DNS 相似。ONS 将 EPC 码转换为一个或多个 URL 地址；通过 URL 地址，可在 EPCIS 服务器上查找关于物品的信息。ONS 提供网络定位，将物品的 EPC 码映射到保存物品信息的 Web 站点。目前 ONS 服务通常用来定位与物品 EPC 码对应的服务器。

通过使用 ONS 基础架构，本地服务器可以查找到 EPC 编码映射的 EPCIS 服务器的 URL 地址。ONS 解析 EPCIS 服务器地址的过程如图 1-18 所示。

对象名称解析服务器 ONS 为用户发起的 EPC 检索请求提供 EPCIS 服务器的地址。从概念上说，ONS 服务的输入就是一个电子产品编码的查询请求，输出则

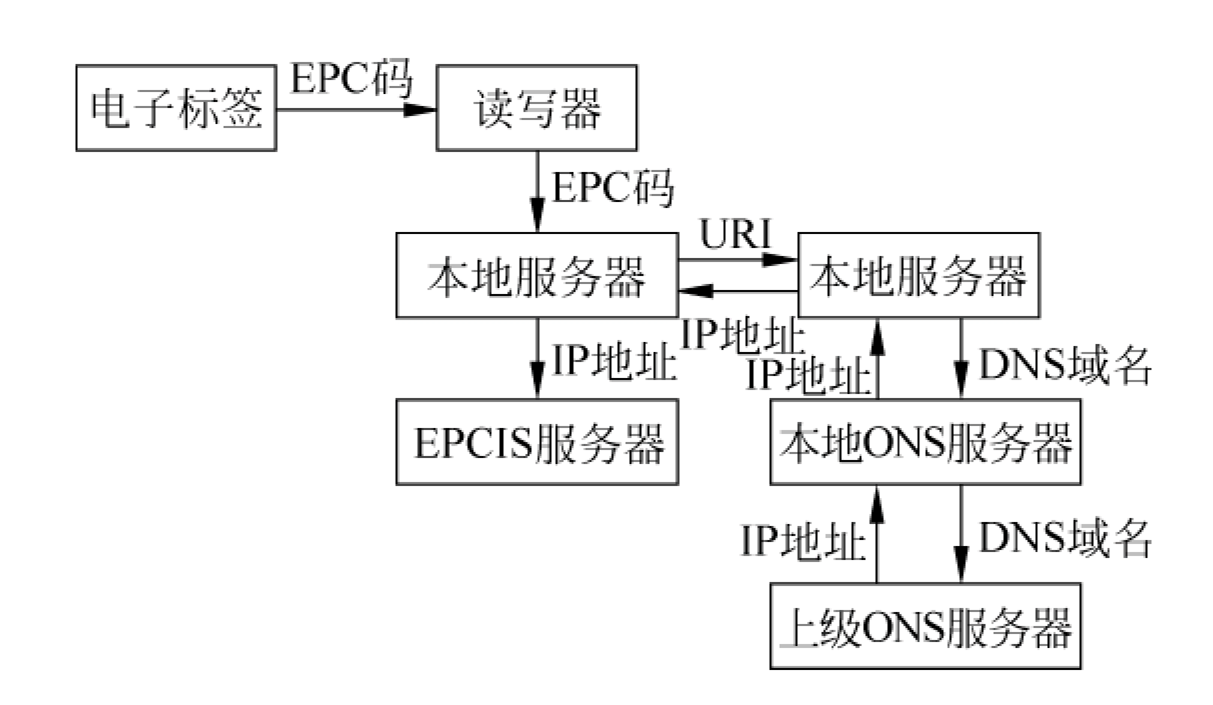


图 1-18 典型的 ONS 查询过程

是所要查找的 EPCIS 服务器的 URL 地址。与 DNS 类似，ONS 也被设计为分级分布的架构，由 ONS 根服务器和本地 ONS 服务器组成。在物联网的应用背景下，与 DNS 相比，ONS 将处理更多的请求。同时，ONS 服务器也需要有冗余性，例如，当一台本地 ONS 服务器崩溃时，ONS 根服务器将能够通过备份找到另一台本地根服务器的 URL 地址，引导完成搜索请求。

当用户希望在 EPCgolbal 网络中的某个位置定位一个 EPCIS 服务时，其请求首先发送到 ONS 根服务器上；ONS 根服务器在根数据表中对该电子产品编码中的 EPC 管理者代码进行解析和识别，并提取该 EPC 管理者所在的本地 ONS 服务器地址，将请求转发至该本地 ONS 服务器；本地 ONS 服务器接收到请求后，进一步在本地数据表中解析 EPCIS 服务器的地址，再将请求转发至该 EPCIS 服务器；EPCIS 服务器最后根据请求的内容提供搜索结果，并返回至发起请求的位置。

4）实体标记语言 PML

PML 是一种用于描述物理对象、过程和环境的通用语言，其主要目的是提供通用的标准化词汇表，来描绘 EPC 编码所标示的物体的相关信息，它是基于人们广为接受的可扩展标识语言（eXtensible Markup Language，XML）而发展起来的。PML 提供了一个描述自然物体、过程和环境的标准，并可供工业和商业中的软件开发、数据存储和分析工具之用。PML 服务器由每个产品制造商维护，用于存储其所有商品的 PML 文件。

5）EPC 信息服务

EPCIS 服务是最终用户与 EPCglobal 网络进行数据交换的主要桥梁，EPCIS 服务器上的数据由供应链上下游的企业共享而获得。通过这种共享，企业可以了解商品在整个供应链环节中的信息，而不仅仅局限于本企业内部。

EPCIS 服务器中共存储两大类数据，分别为：

（1）静态数据，即在产品生命周期中不会发生改变的数据，具体又包括物品类

别静态数据(class-level static data,如产品名称、产品类别等)和属性类别静态数据(Instance-level Static Data,如生产日期、批次号、过期日等)。

(2) 业务数据,即在一件商品流通过程中产生或改变的数据,具体又包括如下三种:

- 属性观察值(Instance Observations),大多由四维数组构成:时间(what)、地点(location)、电子产品编码(who)、业务事件(what);
- 数量观察值(quantity observations),大多由五维数组构成:时间(what)、地点(location)、产品类别(who)、数量(how many)、业务事件(what);
- 业务操作观察值(business transaction observations),大多由四维数组构成:时间(what)、电子产品编码(who)、业务事件(what)、业务操作标识(how)。

EPCIS 提供了一个模块化、可扩展的数据和服务的接口,使得 EPC 的相关数据可以在企业内部或者企业之间共享。EPCIS 有两种运行模式:一种是 EPCIS 信息,可被已经激活的 EPCIS 应用程序直接应用;另一种是将 EPCIS 信息存储在资料档案库中,以备今后查询时进行检索。

1.3　射频技术

射频识别(Radio Frequency Identification,RFID)是一种非接触式的自动识别技术,它利用射频信号通过空间耦合实现非接触信息传递并通过所传递的信息达到识别目的的技术。识别工作无须人工干预,可工作于各种恶劣环境。射频识别技术具有体积小、信息量大、寿命长、可读写、保密性好、抗恶劣环境、不受方向和位置影响、识读速度快、识读距离远、可识别高速运动物体、可重复使用等特点,支持快速读写、非可视识别、多目标识别、定位及长期跟踪管理。

RFID 技术应用于物流、制造、消费、军事、贸易、公共信息服务等行业,可大幅提高信息获取与系统效率、降低成本,从而提高应用行业的管理能力和运作效率。同时,RFID 本身已经成为一个新兴的高技术产业群,成为 IT 产业新的增长点。因此,研究 RFID 技术、开发 RFID 应用、发展 RFID 产业,对提升信息化整体水平、促进经济的发展、提高人民生活质量、增强公共安全等方面有深远的意义。

1.3.1　RFID 系统组成

典型的 RFID 应用系统分为三个主要部分:硬件部分、软件部分和 RFID 中间件部分。硬件部分主要包括 RFID 读写器、天线、标签,将 RFID 读写器放在预先设定好的位置,电子标签贴在货物包装箱(或者附着被识别对象上)上,在 RFID 天线

的识读范围即可实现标签数据的读取。软件部分主要是面向用户的应用软件系统。RFID中间件部分的主要作用是把从硬件部分采集的数据，经过提取、解密、过滤、转换、导入应用软件系统，并通过应用系统呈现在程序界面上，供操作用户浏览、查询、选择、修改。

图1-19是一个典型RFID应用结构，包括数据采集层、RFID中间件层和数据应用层。

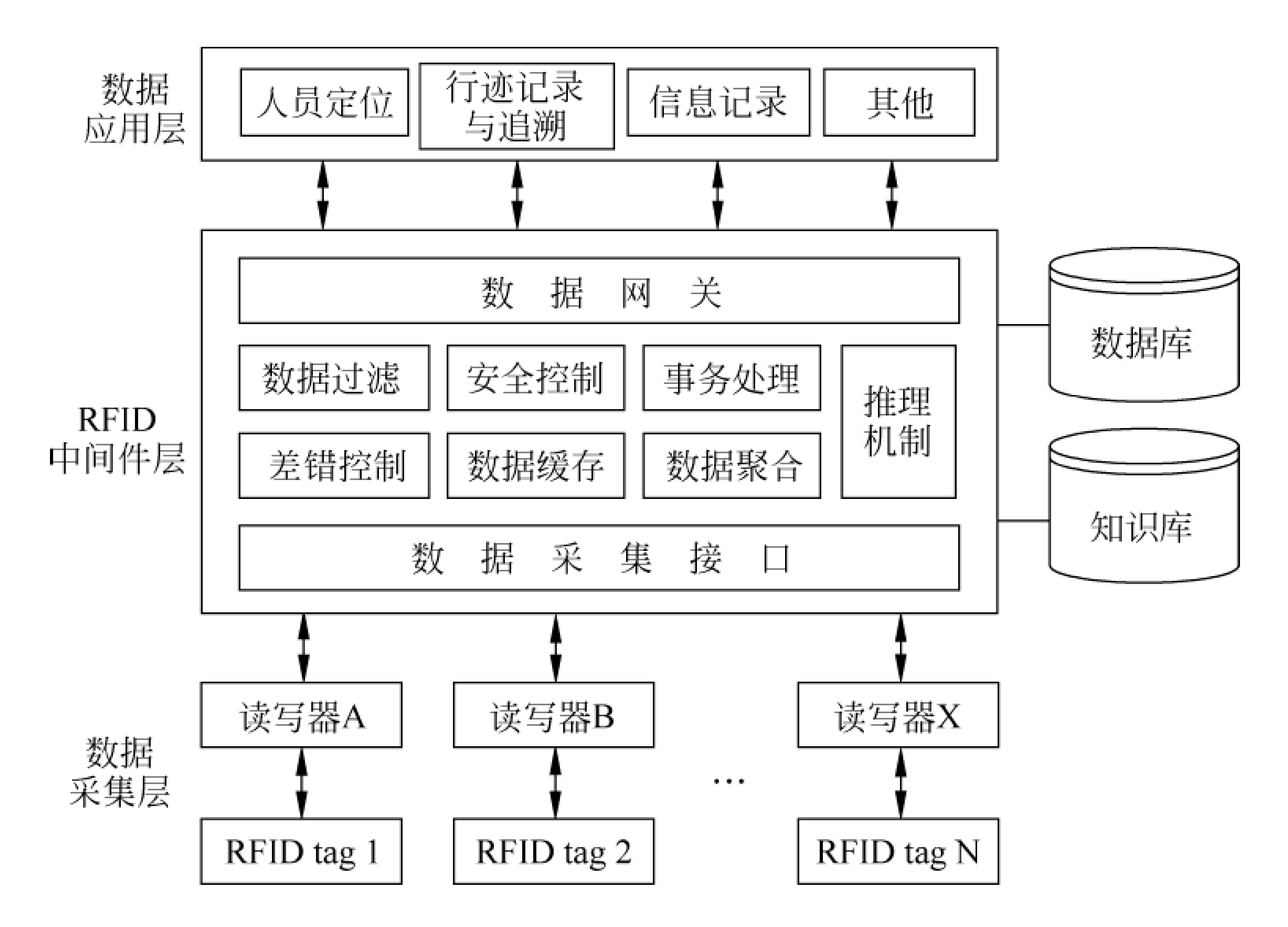


图1-19 RFID应用系统结构图

（1）数据采集层：以读写器和RFID标签为核心，主要涉及读写器与RFID标签之间的通信协议，包括编解码、校验、反碰撞算法等。

（2）RFID中间件层：是连接读写器和各应用软件的纽带，管理标签读写器和应用软件之间的数据流，对数据的安全性、正确性、有效性负责，包括安全控制、数据缓存、过滤、聚合等功能，是整个系统的核心。

同时，也可以建立知识库，根据系统运行中的实际情况，利用知识库中预先设置或自学习的规则来控制电子标签、读写器的工作模式，即向电子标签或读写器发出信令动态改变其工作模式，如休眠状态/工作状态切换、工作频率、发射功率、发送间隔、数据处理方式等。

（3）数据应用层：经过RFID中间件层处理过的数据流，可以提供给不同的应用系统，如人员定位、行迹记录与追溯、人员监控与预警等。

其中，RFID中间件通过两个接口与外界交互：数据采集接口和数据网关。其中，数据采集接口提供多种适配器接口，可以让RFID中间件与不同厂家不同类型的读写器连接；数据网关使RFID中间件与外部不同种类的应用程序或数据库连

接，这些应用程序通常是企业已运行的应用程序和数据库。

在数据网关设计和实现中，一般采用 Web Services 技术，可以使远程的服务器更加快速、安全地得到标签内的信息，及时分析所得数据，而且能够以服务的方式将数据或处理过程提供给其他服务器。RFID 数据采集中间件为应用程序提供了两个服务接口：获得数据集和更新数据集。前者为终端用户提供 RFID 数据，后者为终端用户添加、删除、编辑相关 RFID 数据提供了简单的操作接口。将这两项服务发布到服务注册中心（Universal Description，Discovery，and Integration，UDDI），那么授权用户就可以通过 SOAP 协议绑定这两项服务，实现不同应用系统之间的业务和数据集成。

1.3.2　RFID 硬件部分

RFID 系统中硬件部分通常由标签、读写器两部分组成，如图 1-20 所示。标签中的数据可以由读写器以无线电波的形式非接触地读取。读写器可以将主机的读写命令传送到电子标签，再把从主机发往电子标签的数据加密，将电子标签返回的数据解密后送到主机。主机上的电子标签数据通过 RFID 中间件的 API 被送往 RFID 中间件，经过 RFID 中间件处理后，最终得到应用程序需要的 RFID 相关数据。

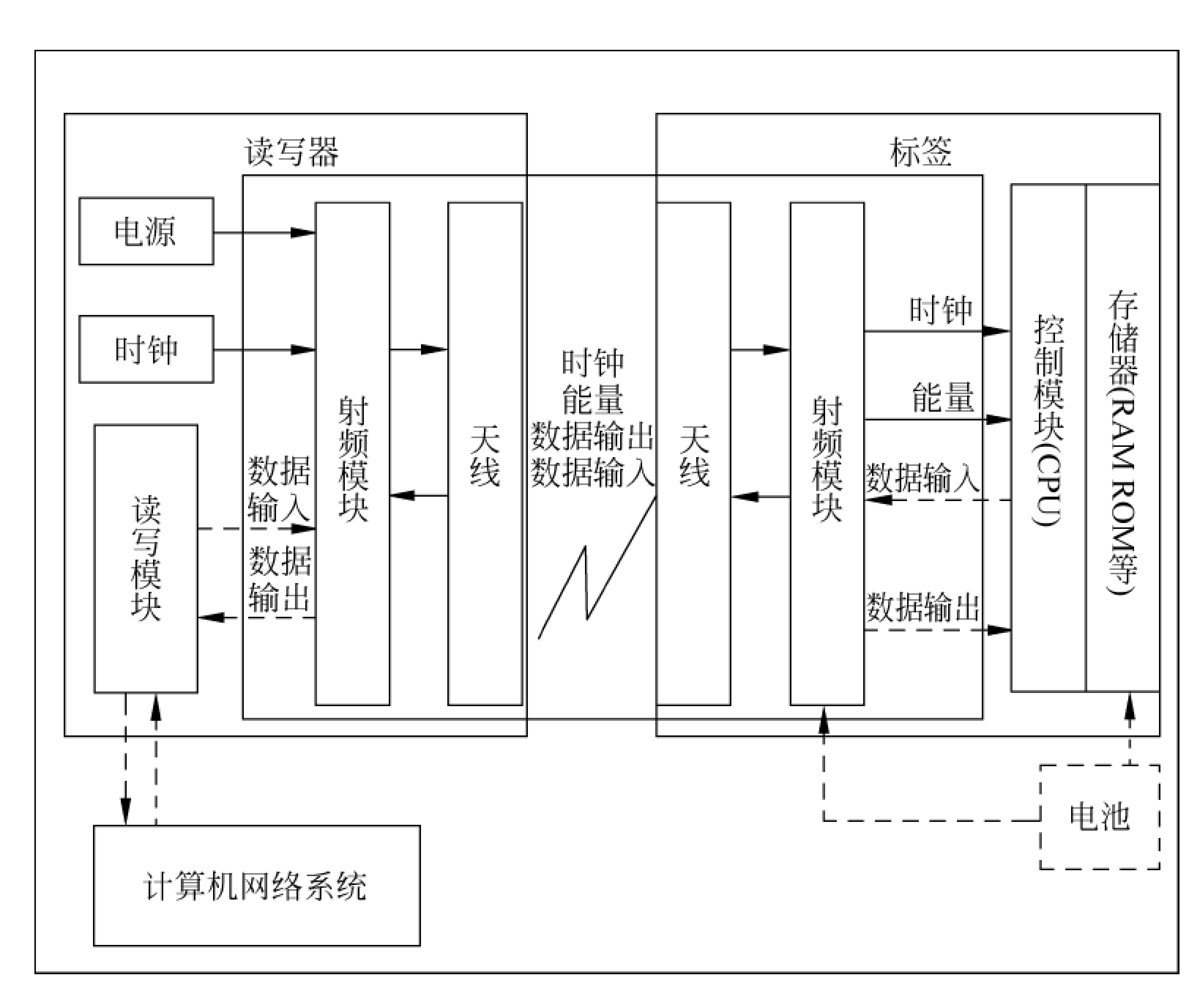


图 1-20　RFID 硬件系统

RFID的基本原理是利用射频信号和空间耦合(电感或电磁耦合)传输特性,实现对被识别物品的自动识别。识别过程中,当电子标签进入读写器所发射电磁波的范围内时,电子标签将读写器所发射的微小电磁波能量存储,进而转换成电路所需的电能,并将存储的信息以电磁波的方式传送给读写器,读写器做出确认及完成后续的控制动作。

1. 电子标签

电子标签是射频识别系统中的重要部件,主要由天线和芯片两部分组成,芯片主要由存储器、控制器、编码器、调制器等组成,其中电源根据标签的发射功率选配,其结构如图1-21所示。

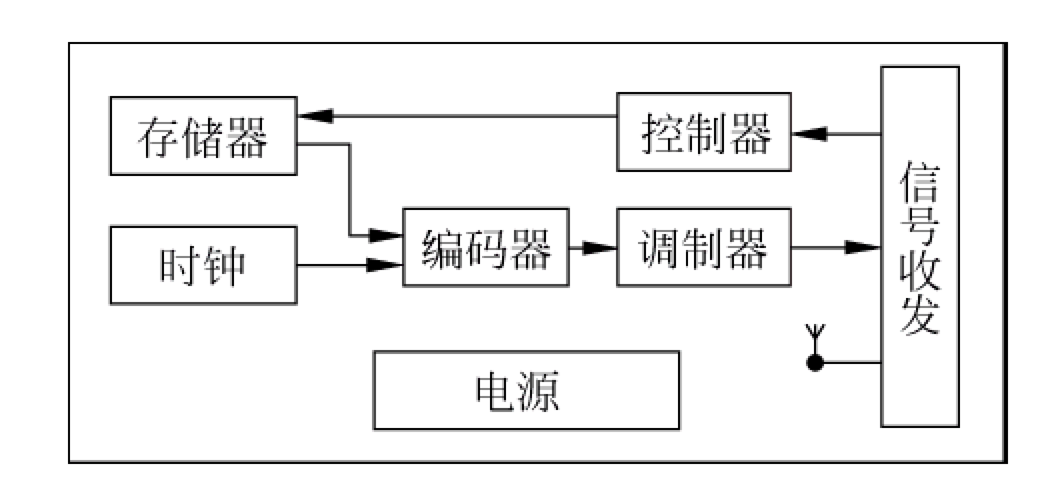


图1-21 电子标签的结构

电子标签根据实际需要可以有不同的样式,每个电子标签都有一个独立且唯一的电子编码,用来区分不同的物品。电子标签按照工作频率的不同可分为低频、高频、超高频;按供电形式可以分为有源标签、无源标签、半无源标签;按调制方式不同可以分为主动式、被动式、半被动式。

射频标签按工作频率分类如下:

1)低频(LF)标签

低频标签工作频率范围30~300kHz,典型的工作频率有125kHz和133kHz。低频标签一般为无源标签,工作能量通过电感耦合(近场)获得,阅读距离小于1m。

典型应用有动物识别、容器识别、工具识别、自动化生产线、精密仪器、电子闭锁防盗等。国际标准有ISO 11784/11785(用于动物识别)、ISO 18000-2(125~135kHz)。

2)高频(HF)标签

高频标签工作频率范围为3~30MHz,典型工作频率为13.56MHz,中高频标签一般也采用无源设置,其工作能量和低频标签一样,也是通过电感耦合(近场)获得,其基本特点与低频标签相似,由于其工作频率的提高,可以选用较高的传输速度,天线设计相对简单,标签一般制成卡片形状。

典型的应用包括:无线IC卡、电子车票、电子身份证、电子闭锁防盗、自动化生产线等。相关的国际标准有ISO 14443、ISO 15693、ISO 18000-3(13.56MHz)等。

3）特高频（UHF）与超高频（SHF）标签

超高频与微波频段的射频标签，简称为微波射频标签。阅读距离一般大于1m，典型情况为4～6m，最大可达10m以上。

各工作频率的用途及特点：

（1）433MHz左右：耦合方式为反向散射耦合（远场），主要用于货物管理及特定场合。该频段电磁波绕射能力强，工作距离较远，但天线尺寸较大，该频段的无线电业务繁杂，容易引起干扰问题。相关的国际标准有ISO 18000-7（433.92MHz）。

（2）800/900MHz频段：我国于近期规划出840～845MHz及920～925MHz频段用于RFID技术，空间耦合方式为反向散射耦合。主要用于商品货物流通。该频段电磁波绕射能力强，最大工作距离可达8m左右，背景电磁噪声小，天线尺寸适中，射频标签易于实现，是全球范围内货物流通领域大规模使用RFID技术的最合适频段。相关的国际标准有ISO 18000-6、EPC GEN2（860～930MHz）。

（3）2.45GHz/5.8GHz频段：空间耦合方式为反向散射耦合（远场）。主要用途为车辆识别和货物流通。该频段电磁波为视距传播，绕射能力差，且相对来讲空间损耗大，因此工作范围小。由于频率高，相对制造成本大，同时该频段为ISM频段，电磁环境复杂，干扰问题在特定场合可能较为突出。相关的国际标准有ISO 18000-4（2.45GHz）、ISO 18000-5（5.8GHz）等。

2. 读写器

读写器是RFID系统的重要组成部分之一，它通过天线向电子标签发送射频调制信号，同时通过天线接收从电子标签返回的载有信息的射频调制信号，经处理后传给中间件或应用系统。典型的读写器由天线模块、射频模块、控制模块和I/O接口等组成，其内部结构如图1-22所示。

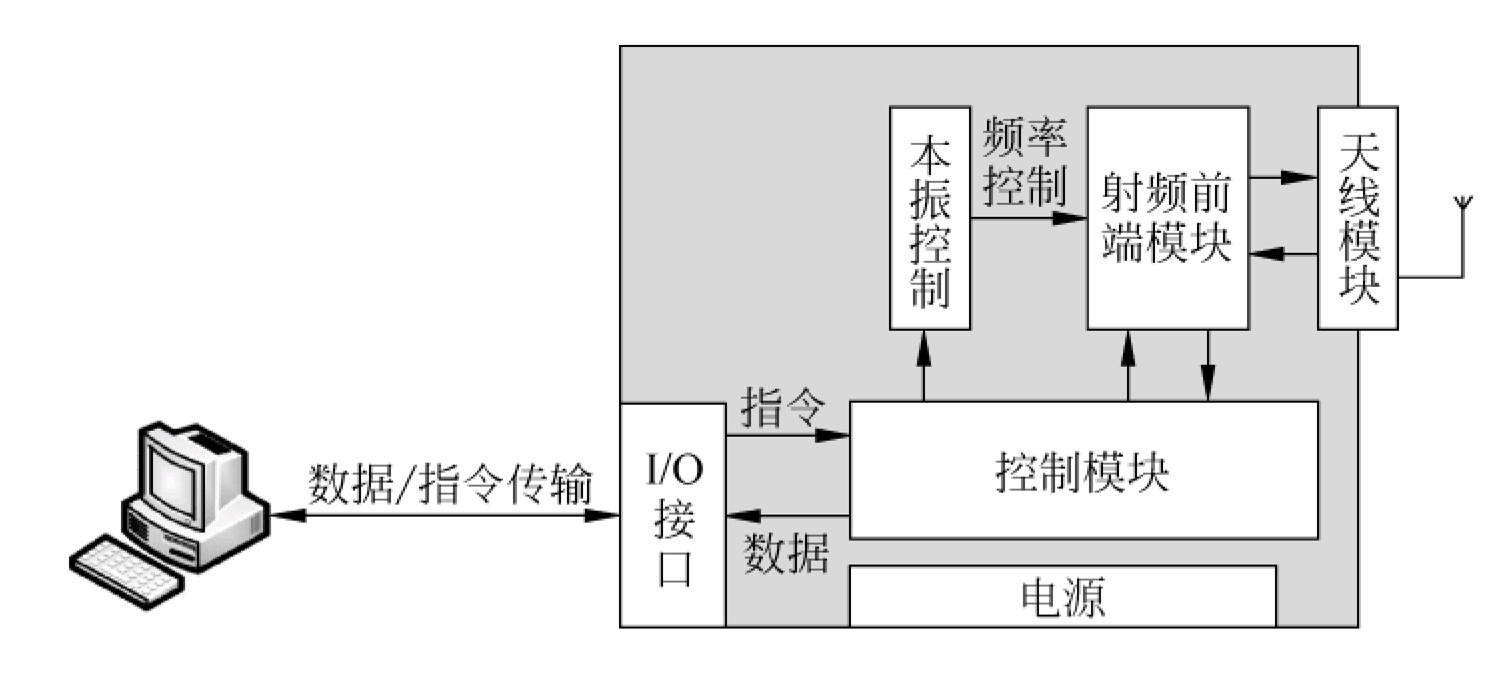


图1-22　RFID读写器的结构

读写器的天线是发射和接收射频载波信号的模块,它主要负责将读写器中的电流信号转换成射频载波信号并发送给电子标签,或者接收标签发送过来的射频载波信号并将其转化为电流信号;射频模块负责接收射频信号和发送包含数据的射频信号;控制模块可调整射频模块发射和接收频率,对发送数据进行编码调制,对接收数据进行解码、校验,并通过数据处理将数据转化为标准格式通过I/O接口发送给应用系统。

1.3.3 RFID 中间件

1. RFID 中间件的分类与特点

RFID 中间件扮演着 RFID 标签和应用程序之间的中介角色,如图 1-23 所示。

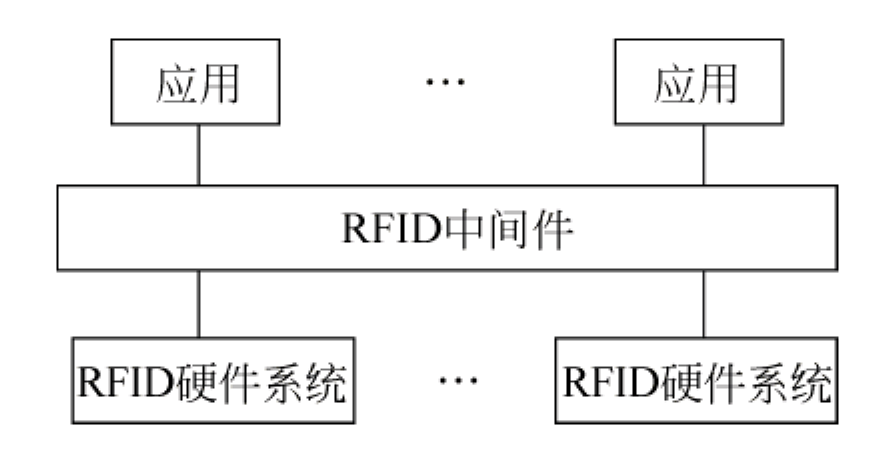


图 1-23 RFID 应用系统

应用程序端通过使用中间件所提供一组通用的应用程序接口(API),连到 RFID 硬件系统。在硬件系统中,读写器读取 RFID 标签数据,这些 RFID 标签数据经过 RFID 中间件的缓存、过滤等诸多操作后提供给应用程序。这样一来,即使存储 RFID 相关数据的数据库软件发生变化,或者后端应用程序增加或改由其他软件取代,或者读写 RFID 读写器种类增加时,应用端不需修改也能处理,避免了多对多连接的维护复杂性问题。

RFID 中间件是一种面向消息的中间件(Message-Oriented Middleware, MOM),RFID 相关数据(Information)是以消息(Message)的形式,从一个程序以异步(Asynchronous)的方式传送到另一个或多个程序。RFID 中间件包含的功能不仅是传递(Passing)信息,还包括安全性、错误恢复、解译数据、数据缓存、数据广播、定位网络资源等高级服务。

一般来说,RFID 中间件具有如下特点:

(1) 独立于架构。

RFID 中间件独立并介于 RFID 读写器与后端应用程序之间,并且能够与多个 RFID 读写器以及多个后端应用程序连接,以减轻架构与维护的复杂性。

(2) 数据流。

RFID 系统的主要目的在于将实体对象转换为信息环境下的虚拟对象,因此数据处理是 RFID 中间件最重要的功能。RFID 中间件具有数据的采集、过滤、整合

与传递等特性，以便将正确的对象信息传到后端的应用系统。

（3）处理流。

RFID中间件采用存储转发（Store-and-Forward）的方式来提供消息流，并具有数据流设计与管理的能力。

RFID中间件从构架上分为两种：

（1）以应用程序为中心的中间件。

通过调用RFID读写器生产厂商提供的应用程序接口API，以Hot Code的方式编写特定读写器的适配器。以应用程序为中心的中间件架构只能实现点对点的连接，仅适用于企业内部单一商业应用系统。

（2）以基础架构为中心的中间件。

随着企业应用系统复杂性的增加，为每个应用通过Hot Code的形式编写Adapter是不现实的，同时面对物件标准化等议题，如EPC，企业可以考虑采用厂商提供的标准化的RFID中间件。因此以基础架构为中心的RFID中间件应运而生。

2. 基于SOA的RFID中间件解决方案

基于SOA的RFID中间件模型如图1-24所示。读写器和射频电子标签构成RFID硬件系统；RFID中间件是RFID硬件系统和企业应用层之间的纽带，同时通过连接ONS服务器和PML服务器，可以在全球范围内形成一种“新式网络”；应用层接收RFID中间件的相关RFID信息数据，是RFID数据的后端应用部分。

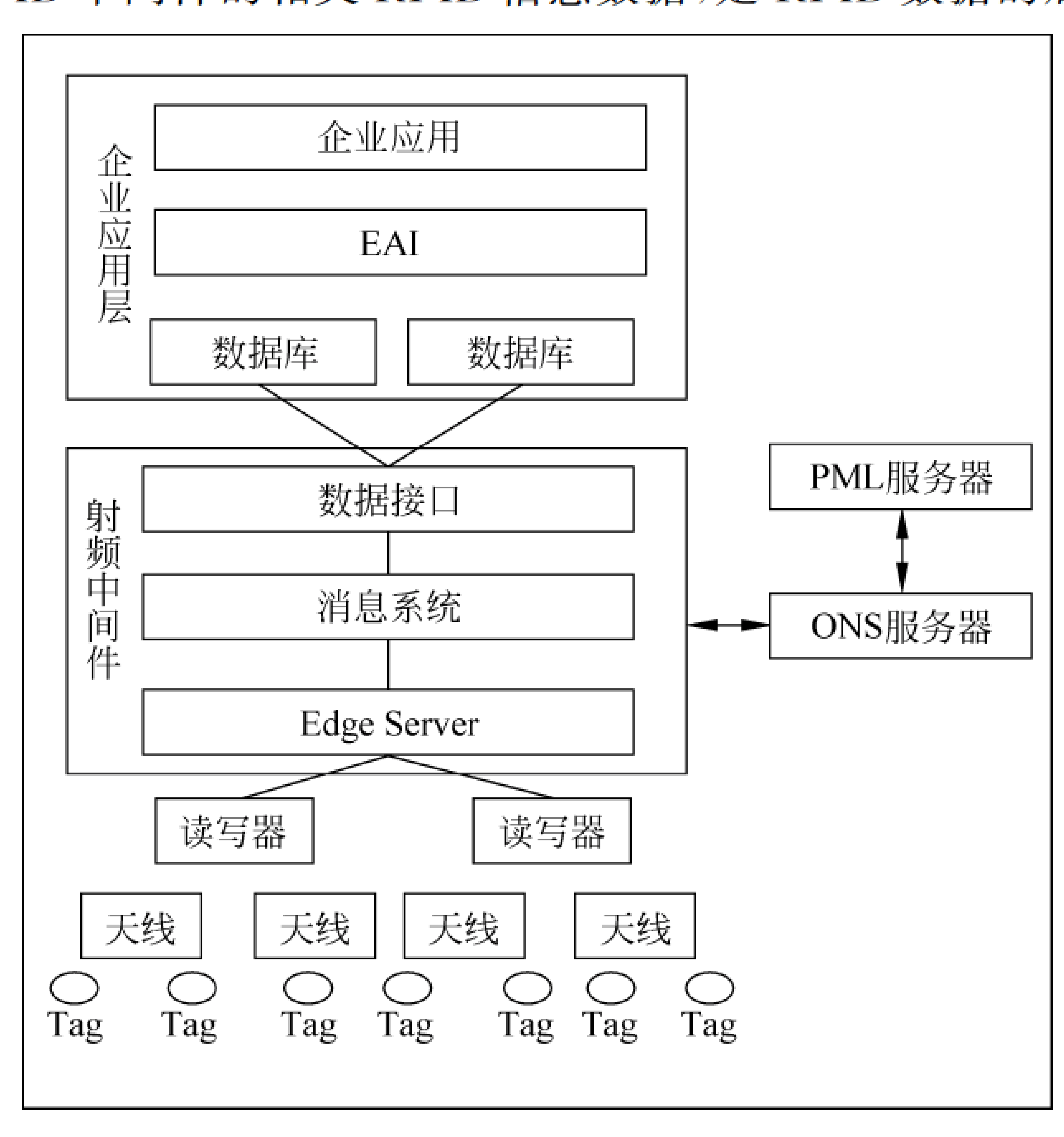


图1-24　基于SOA的RFID中间件架构

RFID 中间件分为 Edge Server、消息系统和数据接口三个层次，这三个层次有着明确的功能划分，描述如下：

1） Edge Server

“Edge Server”即边缘服务器，位于 RFID 中间件的最底层，直接和读写器交互，主要功能包括：

（1） 对电子标签中的数据进行采集；

（2） 对于来自不同类型的读写器的数据进行适配处理，得到统一的、格式化的数据；

（3） 对适配处理后的 RFID 相关数据进行校验；

（4） 将校验无误的 RFID 相关数据按照用户定义的协议进行消息包的封装，并将消息包发送到消息系统。

Edge Server 的结构图描述如图 1-25 所示。其中，读写器接口完成数据采集和数据适配处理工作；数据校验工作单元完成对来自读写器接口的数据的校验；数据封包工作单元对来自校验工作单元的 RFID 数据，依据数据内容将这些 RFID 数据打包成不同的消息，并传递到 RFID 中间件的下一个功能单元——消息系统中。

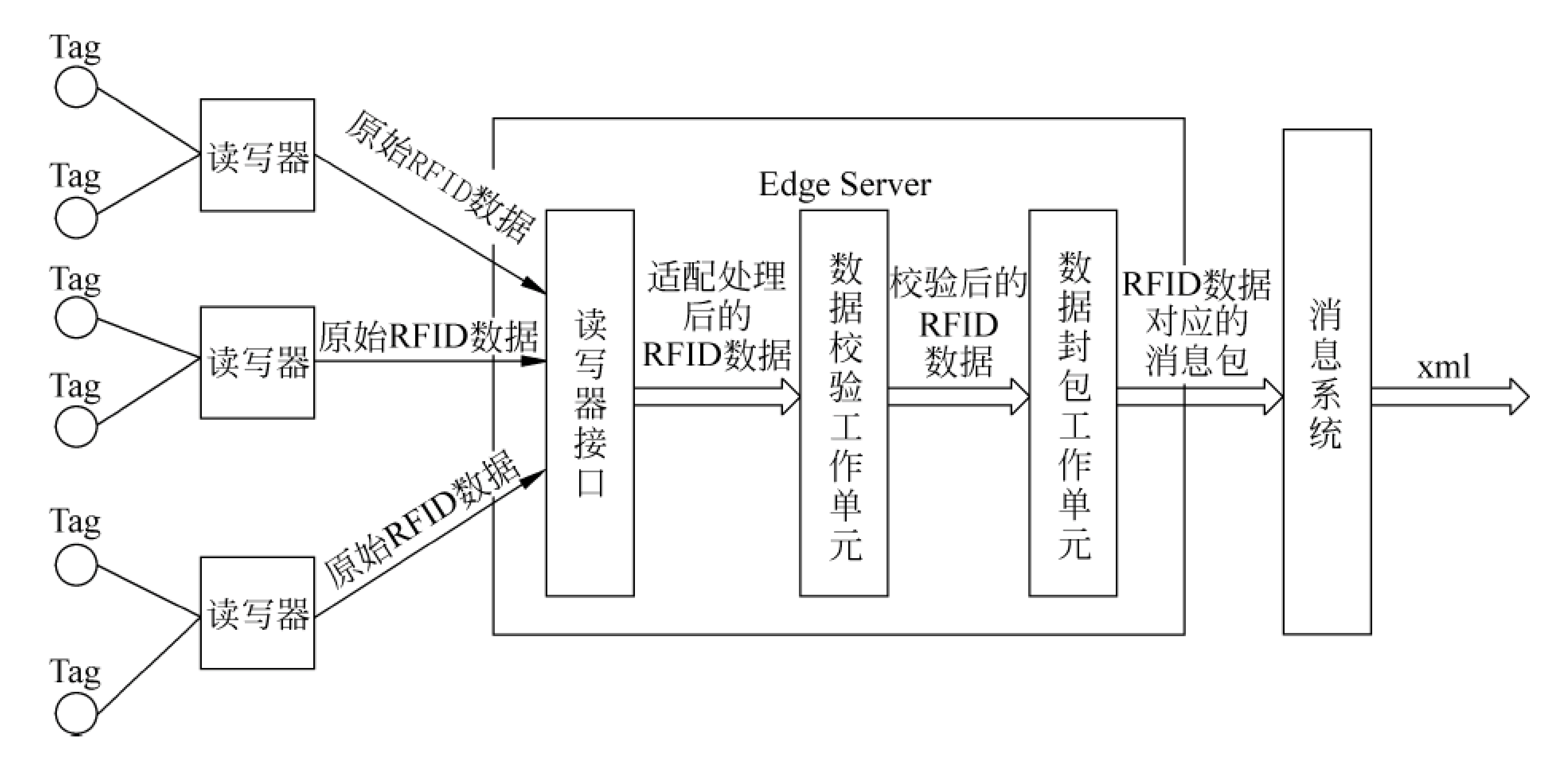


图 1-25 边缘服务器

2） 消息系统

在 RFID 系统中，一方面是各种应用程序以不同的方式频繁地从 RFID 系统中取得数据；另一方面却是有限的网络带宽。因此，需要采用消息系统解决这一问题。

如图 1-26 所示的消息系统位于 RFID 中间件的中心层。Edge Server 产生事件，并将事件传递到消息系统中，由消息系统决定如何将事件数据传递到相关的应

用系统。处理过程描述如下：消息系统首先在消息服务器上缓存来自 Edge Server 上的各种消息，然后依据消息内容将这些消息分类整合，使得同类消息位于相同的消息队列中，最后将分好类的消息分别存储成相应的 xml 临时文件，这些临时 xml 文件最终会被送往数据接口做进一步的处理。

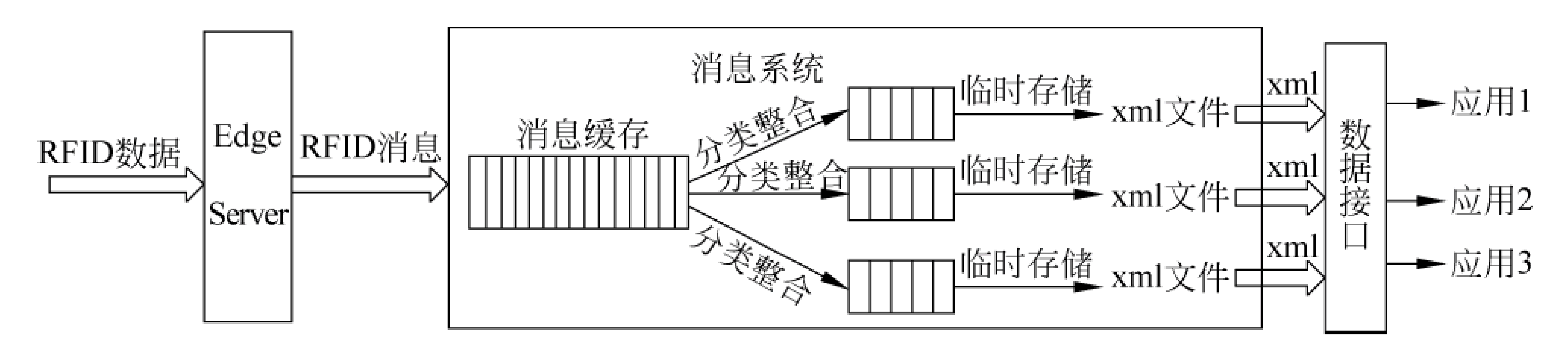


图 1-26　消息系统

在这种模式下，读写器不必关心哪个应用系统需要什么数据；同时，应用程序也不需要维护与各个读写器之间的网络通道，仅需要将需求发送到消息系统即可。消息系统具有如下功能：

(1) 数据缓存功能。由于同一时间来自 Edge Server 的数据量是巨大的，如果将这些数据直接进行后台数据库的入库操作，这对于数据库的接受能力是一个巨大的考验。因此在入库前将这些数据通过消息队列先进行缓存，有助于缓解数据库压力。

(2) 基于内容的路由功能。由读写器获取的全部原始数据，在多数情况下能够用到的仅仅是其中的一部分。例如设置在仓库门口的读写器读取了货物消息和托盘消息，但是业务管理系统只需要货物消息，固定资产管理系统需要托盘消息。这就需要消息系统具有依据事件消息内容来决定消息传递方向的功能，否则将导致应用程序不得不自己实现部分的过滤工作。

(3) 数据分类存储功能。有些应用(如物流分拣系统或销售系统)需要实时得到读取的标签信息，所以消息系统几乎不需要存储这些标签数据。而有些系统则需要得到批量 RFID 标签数据，并从中选取有价值的 RFID 事件信息，这就要求消息系统应该提供数据存储功能。

3) 数据接口

来自消息系统的数据最终是归好类的 xml 磁盘文件。同一类型的数据以 xml 文件的形式保存，并供相应的一个或多个应用程序使用。而数据接口主要是对这些数据进行过滤、入库操作，并能提供访问相应数据库的服务接口。数据接口的结构图如图 1-27 所示，其具体工作如下：

(1) 将存放在磁盘上的 xml 文件进行批量入库操作，也就是说当 xml 中数据量达到一定数量时，启动数据入库功能模块，将 xml 中的数据移植到各种流行数据

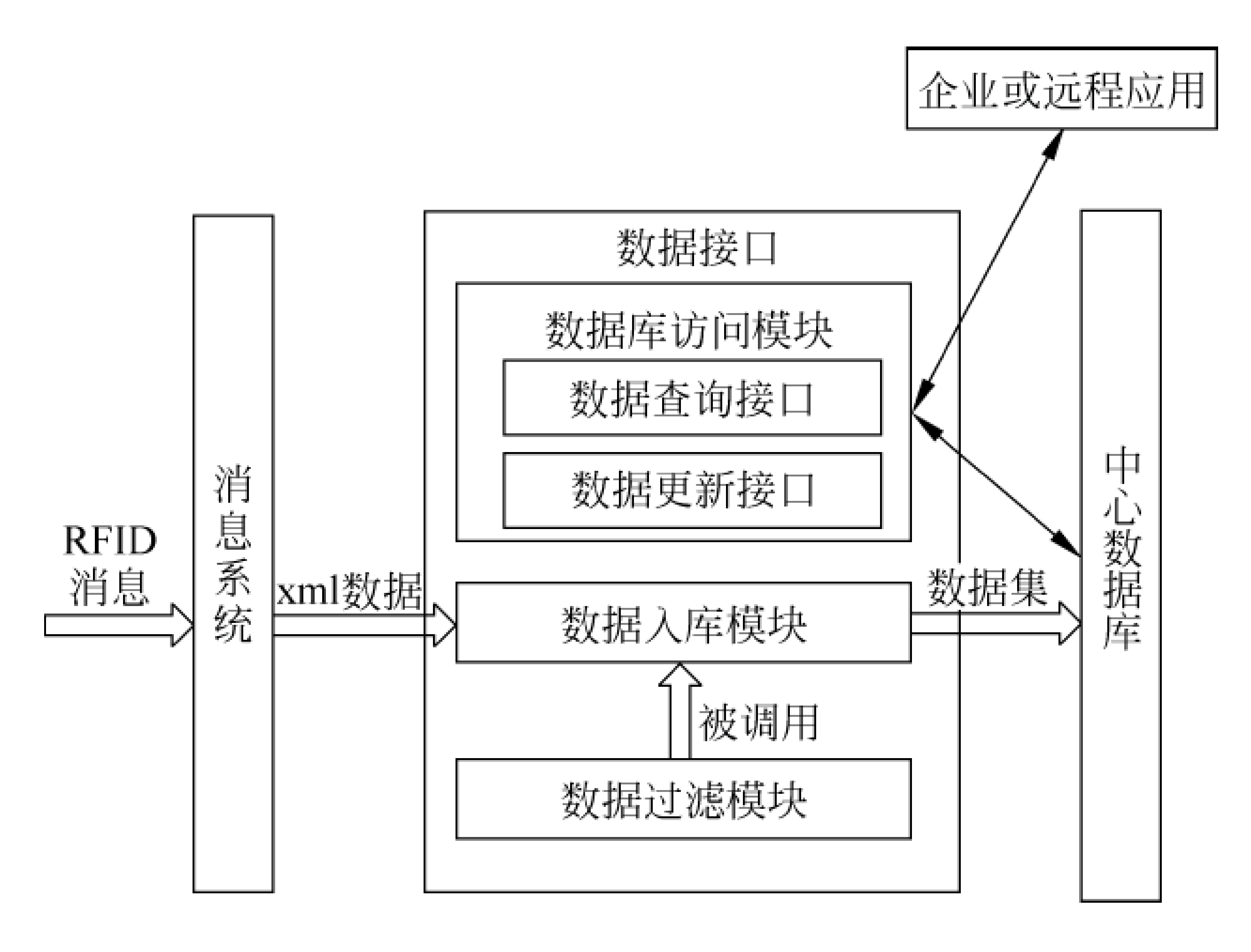


图 1-27　数据接口

库中，如 SQL Server、Oracle 等；

(2) 在 xml 数据入库前将重复的 xml 数据过滤掉；

(3) 为企业内部和企业外部访问数据库提供 Web Services 接口。

1.4　生物识别技术

1.4.1　生物识别技术概述

20 世纪 80 年代末 90 年代初，随着信息安全重要性的日益突出，生物特征认证技术研究开始成为一个研究热点，一些以生物特征认证为核心的技术和产品不断涌现。

美国“911”事件之后，如何通过高科技手段来高速准确地鉴定和认证个人身份成为各国政府和公众极为关注的一个话题，而在所有需要进行身份识别的地方，都可以应用生物特征认证技术，因此生物特征认证技术受到了前所未有的重视。生物特征身份认证技术(Biometrics)是指通过计算机与光学、声学、生物传感器和生物统计学等技术结合，利用人体固有的生物特性(如指纹、人脸、虹膜等)和行为特征(如笔迹、语音、步态等)来进行个人身份的认证。生物特征身份认证技术具有不会遗忘、不易伪造或被盗、随身携带和随时随地可用等优点，比传统的身份认证方法更加安全、保密、方便。目前，比较成熟和最具有应用前景的几种生物特征识别技术包括指纹、虹膜、人脸、视网膜、手形、手部脉络、语音以及签名等。

1.4.2 基于指纹的身份认证

英国的亨利·福尔茨(Henry Faulds)在19世纪末20世纪初促进了指纹的研究和发展,1880年他首先发现了至今仍被承认的指纹的两个重要特征:一是任何两个不同手指的指纹脊线的式样(Ridge Pattern)不同,二是指纹脊线的式样在人的一生中不会改变。这一发现奠定了现代指纹认证技术的理论基础,也使得指纹认证在犯罪鉴定中得到应用。在20世纪初期,司法部门已经正式采用指纹作为有效的身份标记,一些指纹认证机构建立了世界范围的罪犯指纹档案库。

20世纪80年代,个人计算机、光学扫描这两项技术的革新,使得它们作为指纹取像的工具成为现实,从而使指纹认证可以应用于其他需要进行身份认证的领域;20世纪90年代后期,电容传感器等廉价取像设备的引入及其飞速发展,使指纹认证系统的体积和价格得以大幅度减低,加上较先进的认证算法的提出,为个人身份认证应用的增长提供了较好的技术基础。

相对于其他生物特征,指纹有如下两个突出的优点:

(1) 稳定性。指纹具有很强的相对稳定性。从胎儿六个月指纹形成到年老,指纹纹线类型、结构、统计特征的总体分布始终不会有明显变化。即使手指皮肤受伤,只要不伤及真皮层,伤愈后纹线仍能恢复原状。

(2) 唯一性。指纹具有明显的唯一性,至今仍找不出两个指纹完全相同的人。指纹的唯一性是由亨利·福尔茨提出的,这是他在日本采样研究的一个结果。用数学方法证明指纹各不相同的是法国巴黎大学教授勃太柴,他在1910年证明了此特性。由于皮肤表皮上的纹路是在胎儿六个月时形成的,所以即使是同卵双胞胎的指纹也是不同的。不仅是人与人之间,就是同一个人的十指指纹也有明显的区别。

指纹身份认证一般都要经过图像获取、特征提取和指纹匹配三个过程。图像获取是指通过设备获取手指指纹并转化为数字图像的过程,特征提取是从图像中提取指纹特征,指纹匹配一般是指匹配指纹的特征。

指纹认证系统分为两个阶段:注册阶段和认证阶段。在注册阶段,用户需要输入用户名,同时采集指纹(将手指放在指纹采集设备的传感器窗口上,以便采集设备采集指纹),自动指纹认证系统的特征提取模块会从输入的指纹中提取特征,最后系统将用户名和提取的指纹特征一起保存到数据库中。保存在数据库中的指纹特征会在认证阶段被系统读取,用来与输入指纹的特征匹配。具体过程如图1-28所示。

1.4.3 基于虹膜的身份认证

最早的虹膜身份认证可以追溯到1885年,巴黎刑事监狱利用生物特征来认证

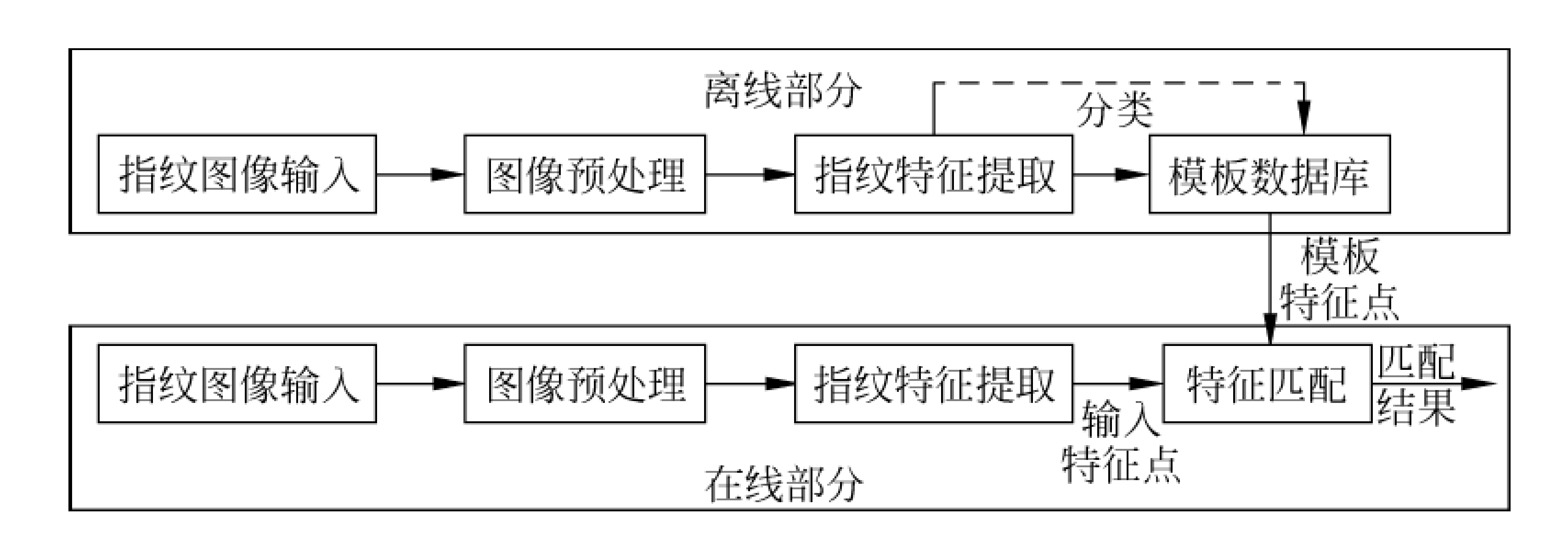


图 1-28 指纹识别系统框架

囚犯。当时用耳朵的大小、脚的长度和虹膜等生物特征区分同一监狱中的不同囚犯。1987 年，美国的眼科专家 Leonara Flom 和 Aran Safir 在他们的专利文献中指出人眼虹膜纹理特征具有独特性，即便是同一个人，其左眼和右眼的虹膜纹理特征也不相同，而且虹膜发育完全后，其纹理将终生不变。这是虹膜可以作为一种身份认证手段的物理基础。

虹膜认证是现代成熟起来的一种利用瞳孔和巩膜之间的环状区域纹理进行身份鉴别的生物特征认证技术。虹膜位于巩膜和瞳孔之间，包含了丰富的纹理信息，不同种族的人虹膜颜色不同。

虹膜身份认证技术的基本步骤为：

(1) 图像采集，用于获取虹膜图像；

(2) 预处理，进行虹膜内外边缘定位、归一化和图像增强等；

(3) 特征提取，得到虹膜纹理的特征编码；

(4) 特征匹配，将提取的虹膜特征编码与特征模板进行匹配以区分不同的虹膜。

整个步骤如图 1-29 所示。

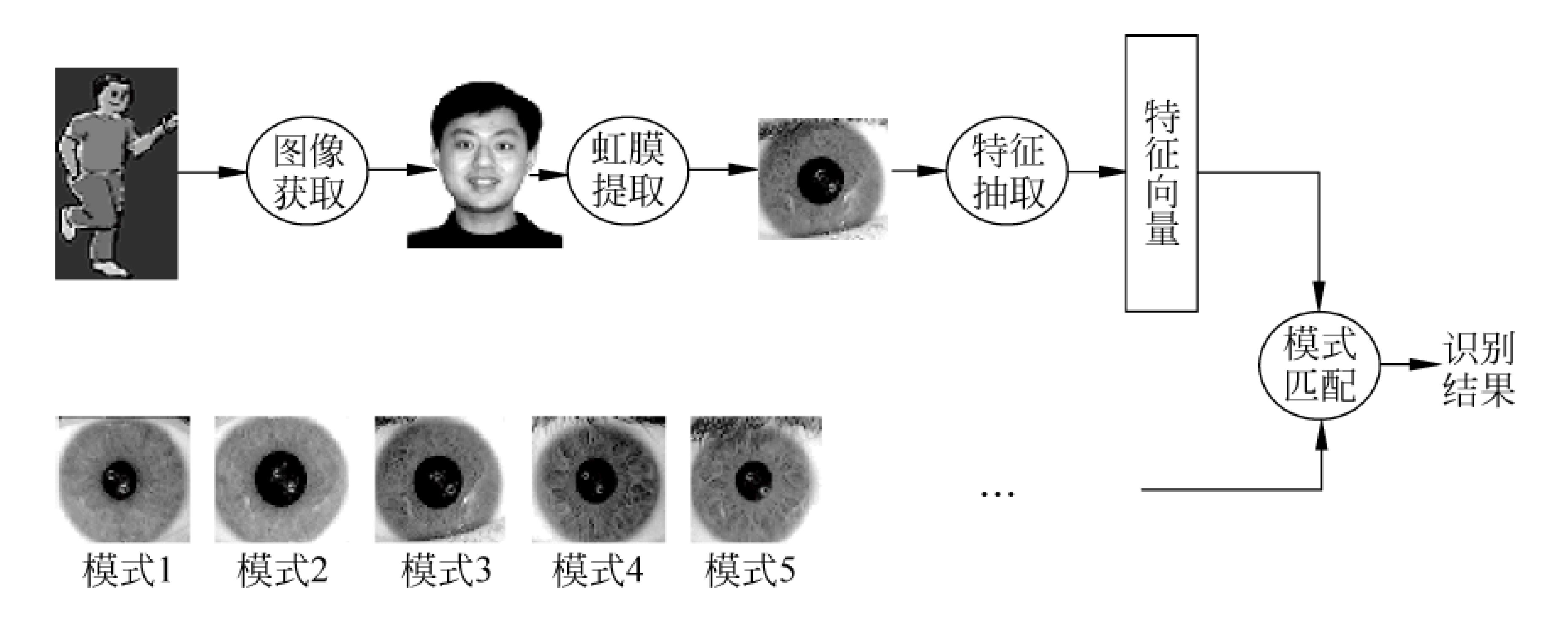


图 1-29 自动虹膜识别系统示意图

1.4.4　基于人脸的身份认证

近年来,人脸认证研究得到了诸多研究人员的青睐,涌现出了诸多技术方法。尤其是1990年以来,人脸认证更得到了长足的发展。

与其他生物特征认证技术相比,人脸认证在应用方面具有独到的技术优势:

(1) 可以隐蔽操作,尤其适用于安全监控。

(2) 非接触式采集,没有侵犯性,容易被接受。

(3) 具有方便、快捷、强大的事后追踪能力。基于面相的身份认证系统可以在事件发生的同时记录并保存当事人的面相,从而可以确保系统具有良好的事后追踪能力。

(4) 图像采集设备成本低,使用中低档摄像头、数码相机、数码摄像机和照片扫描仪等设备即可完成图像采集工作。

(5) 更符合人类的认证习惯,可交互性强。一般用户对于指纹、虹膜等认证系统往往无能为力。对人脸来说,授权用户的交互和配合可以大大提高系统的可靠性和可用性。

人脸是人类一个非常独特的视觉特征,人脸认证问题可以看作计算机视觉领域里一个具有代表性、非常典型的问题。比起其他生物认证方式,人脸认证是一种更直接、更方便、更友好、更容易被人们接受的非侵犯性的认证方法,所以近年来以人脸为特征的生物认证技术发展十分迅速。人脸认证的主要内容概括为以下几个方面。

1. 人脸表征

人脸表征是提取人脸的特征,是将现实空间的图像映射到机器空间的过程。人脸的表征具有多样性和唯一性,可以保证人脸图像的准确描述和识别。

人脸图像信息数据量巨大,为了提高检测和识别的运算速度,提高图像传输和匹配检索速度,必须对图像进行数据压缩,降低向量维数,即用尽可能少的数据表示尽可能多的信息。

2. 人脸检测

人脸检测的任务是从一幅图像中判断是否存在人脸,找出人脸所在位置与其所占区域。检测的准确度受光照条件、成像器材质量、遮挡、人脸大小、角度、表情等多方面因素的影响。

3. 人脸跟踪

人脸跟踪是根据已定位出的人脸,在后续图像帧中持续地跟踪该人脸的运动。其难点主要是需要花费较大的计算量,并会受到光照变化的影响。

4. 人脸识别

人脸识别的任务是识别或确认当前人脸，在人脸位置确定后可以对人脸进行认知，即将人脸与数据库中的人脸进行比较，得出有关身份方面的信息。

5. 表情/行为分析

表情/行为分析是让计算机感知使用者的表情变化，分析理解人的情绪，如快乐、愤怒、忧伤、赞同或否认等，从而做出进一步的反应。

综上所述，一个完整的人脸认证系统的框架如图 1-30 所示。其中输入分为静态图像和动态视频，先进行人脸检测，确定图像中人脸的位置，选取合适姿态的人脸进行认证；之后是器官定位和人脸归一化，调整人脸方向，统一人脸大小，以利于与数据库中的人脸进行比对，最终识别/确认输入的到底是哪张人脸。

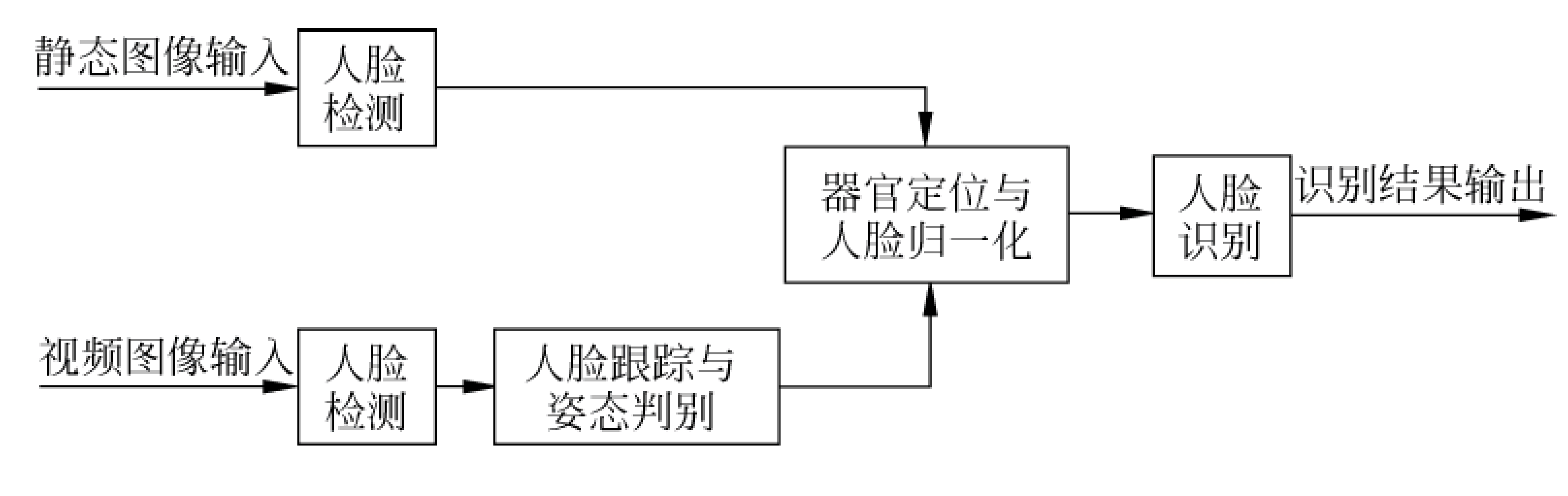


图 1-30 人脸识别基本框架

1.4.5 基于其他生物特征的身份认证

1. 人脸温谱图认证

人脸温谱图是利用红外传感器获得的人脸热辐射图像。研究表明，每个人的人脸温谱图不但具有唯一性，而且具有较高的稳定性和持久不变性，即使通过外科整形手术也很难改变，因此可以用来认证个人身份。人脸温谱图认证技术具有非接触性和非侵犯性、红外传感器可在弱光或没有光照的情况下捕获人脸温谱图等优点，其缺点是容易受周围环境温度影响较大，并且红外传感器价格较高，从而限制了它的广泛使用。

2. 视网膜认证

视网膜认证是通过视网膜读取器感知人眼后面的视网膜脉络模式而进行识别的。该认证方法要求被识别者的眼睛对着一个目镜往里看，这时从眼睛后部反射出来的光线用于捕捉视网膜脉络模式。

视网膜认证的优点在于具有很高的识别率。但是视网膜技术可能会给使用者带来健康的损坏，而且设备投入较为昂贵，识别过程的要求也高。

3. 手部脉络模式认证

手部脉络模式是指人的手部血管脉络的纹路和结构模式。每个人的手部脉络模式是各不相同的,即它具有唯一性,因此可用于身份的认证。手部脉络模式认证技术的优点是:手部脉络模式的数字图像可以用红外摄像机拍摄,而且易于从背景中分离。缺点是:人的手部脉络模式稳定性较差,会随着年龄的增长而发生改变;基于手部脉络模式的身份识别系统尺寸较大,也给使用带来不便。

4. 手形认证

手形是人手独有的形状特征,这些特征包括手的外部轮廓、手指的长度和大小等。手形认证系统的优点是:简单、费用低;容易被人们接受;手形图像采集容易,对健康没有影响。缺点是:手形的大小会随着年龄的增长而发生改变,因此手形认证系统要定期更新数据库,这给系统的设计和使用带来许多不便;劳动和受伤等情况会使手形发生变化,给系统的正常判断带来困难。

5. 签名认证

签名作为身份认证的手段已经有几百年的历史了。人们已经熟悉了在银行的格式表单中或在法律上有效的文件上签名作为自己身份的标志。作为一种行为认证技术,签名认证是通过测量图像本身以及整个签名的动作,即在每个字母以及字母之间的不同速度、顺序和压力而获得签名特征的唯一性。签名认证容易被大众接受,而且是一种公认的身份认证技术。但是由于经验的增长、性情的变化和心情的改变,签名也会随之改变。另外,还要注意所使用的签字笔的磨损程度等。所有这些都增加了签名认证系统的识别难度,降低了认证结果可靠性。

6. 语音认证

语音认证本质上也是一个模式识别问题。在语音认证中,需要说话人讲一句或几句测试短句,对它们进行某些测量,然后计算量度矢量与存储的参考矢量之间的一个(或多个)距离函数,最后进行模式识别。语音认证技术的优点是语音信号获取方便,其缺点是在有噪声的环境中或感冒时语音会受到影响使识别率下降,这些使得其应用范围受到很大的限制。

除了上面提到的生物特征识别技术外,还有许多已经研究或正在研究的生物特征识别技术。例如,耳朵形状、嘴唇形状、人体气味、步态等。

1.5　位置服务技术

近年来,随着无线局域网的发展,对室内外环境中人员和物体的追踪定位引起了研究者的广泛关注,基于位置的服务(Location Based Services,LBS)越来越受到

人们的关注。

基于位置服务的最早应用是20世纪70年代美国的911电话定位；之后，随着GPS(Global Positioning System，全球定位系统)的出现，导航成为基于位置服务的最典型的应用。如今，随着定位手段和技术的发展、移动计算设备的逐渐普及以及移动网络通信技术的发展，使得基于位置服务的普及成为可能。

在社区矫正管理工作中，基于位置的服务器系统可以帮助司法行政机构对矫正人员的日常活动情况进行监测和管理。在社区矫正中，通过无线定位技术可以实现矫正人员的实时定位查询、自动跟踪及历史轨迹查询和回放，超越规定区域自动报警等。

1.5.1 位置服务系统

基于位置的服务的硬件环境由通信平台、定位平台和地图平台三部分构成(如图1-31所示)。通信平台利用各种有线、无线网络传输移动对象实时位置，为用户访问位置服务平台的基础网络链路；定位平台利用GPS或北斗卫星定位系统，以及地面无线定位技术实时确定移动对象的位置；地图平台负责地理空间信息管理、可视化、信息查询与实时引导等。数据采集与处理是基于位置的服务的重要环节，也是其服务质量的保障基础。基于位置的服务所涉及的信息主要包括位置信息、地理空间信息等。位置信息记录移动目标所处的空间位置；地理空间信息描述移动目标所处的地理环境，如道路网络、建筑物、三维街景等。

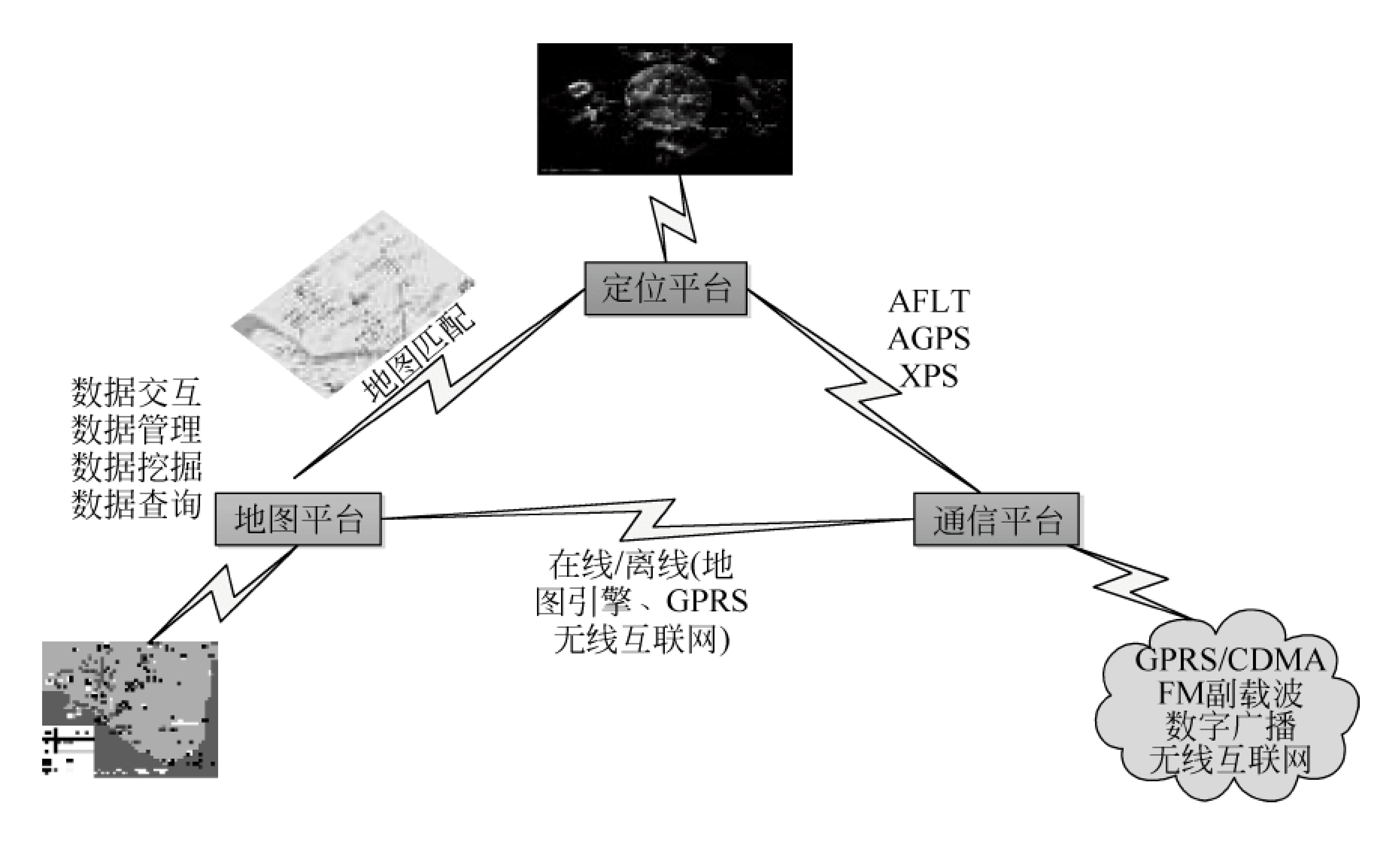


图1-31 基于位置的服务平台结构

1.5.2　室内定位方法分类

作为 LBS 的核心技术之一，无线定位是其重要的方法之一。无线定位是指利用无线电波信号的特征参数估计特定物体在某种参考系中的坐标位置。

1. 基于位置感知技术的分类

室内定位技术的分类方法比较多，基于位置感知技术可以将室内定位技术分为如下三种基本类型：

（1）基于临近关系的定位技术，根据待定位物体与一个或多个已知位置的临近关系来定位；

（2）基于三角关系的定位技术，根据采用三角形的几何特性计算物体的位置；

（3）基于场景分析的定位技术，利用从某一优势位置观察到的场景中的特征信息，来估计观测者的位置或者场景中某一物体的位置。

2. 基于信号测量技术的分类

基于信号测量技术，室内定位技术又可以分为如下类型：

（1）基于 RSS（Received Signal Strength，接收信号强度）的定位，依据接收信号强度与距离的传播模型关系将接收信号强度转换为距离值，再利用三角关系计算位置；

（2）TOA（Time of Arrival，到达时间），测量发送端到接收端的信号传输时间，乘以信号传播速度得到距离值，再利用三角关系计算位置信息；

（3）TDOA（Time Difference of Arrival，到达时间差），测量发送端到不同接收端的信号传输时间差，转换为发送端到不同接收端的距离差值，利用几何关系计算位置信息；

（4）AOA（Arrival of Angle，到达角度），测量信号到达的角度，利用几何关系来确定目标的位置；

（5）基于 Cell-ID（Cell Identification，基站）的定位，根据目标所处小区标识（Cell-ID），利用临近关系来确定目标的位置；

（6）基于 BER（Bit Error Rate，误码率）测量的定位，测量接收端的信号的传输错误率，依据传输错误率与距离的模型关系，将传输错误率转换为距离值，再利用三角关系计算位置信息。

3. 基于传感器类型的分类

根据传感器类型的不同，室内定位技术还可以分为：基于 RFID 系统的定位、基于 WLAN 系统的定位、基于红外线系统的定位、基于超声波系统的定位、基于蓝牙系统的定位、基于超宽带系统的定位、基于 WSN（Wireless Sensor Network，无

线传感器网络）系统的定位等。典型的室内定位系统主要有：基于 RFID 的 SpotON 系统；基于红外线的 Active Badge 系统；基于超声波的 Cricket System 系统和 Active Bat 系统；基于蓝牙的 Tadlys 公司开发的系统；基于超宽带的 Localizers 系统和 Sapphire 系统；基于 WLAN 的 RADAR 系统等。

1）RFID 定位技术

RFID 室内定位分为距离估计和场景分析方法。距离估计主要是运用 RSS、TOA 等参数，根据特定的物理模型或者几何模型对于距离进行估算；场景分析一般是预先建立基于 RSS 参数值的先验分布图，然后根据概率模型、*k*-NN(*k*-nearest neighbor)模型等做出相应的分析，实现定位；这两大类方法要求 RFID 读写设备上有度量特定物理参数的模块，通过收集参数，进行运算实现定位。

2）基于 WLAN 的定位技术

近几年，基于 WLAN 的室内定位技术已成为研究与应用的热点，典型的方法包括基于信号传播模型和基于指纹模型的定位方法。基于传播模型的方法是指利用无线信号在空气中传播时所呈现出的非线性衰减特性，建立信号强度与距离的模型，从而获得位置信息；而基于指纹模型的定位方法主要是指基于大量实际观测数据，采用模式识别的方法进行定位模型训练，然后用获得的定位模型对目标点进行位置估计。后一种方法一般分为离线训练和在线定位两个阶段，离线训练阶段采集所需定位区域各参考节点的信号特征参数，建立位置指纹数据库；在线定位阶段，采集测试地点的信号参数，采用模式匹配算法与指纹数据库中的数据相匹配，得到用户的位置估计。

3）基于 WSN 的定位技术

无线传感器网络（Wireless Sensor Network，WSN）是一种由传感器节点用自组网方式构成的大规模网络，它将通信、嵌入式计算和传感器技术三大技术基础相结合，成为近年出现的比较热门的研究领域之一。无线传感器网络作为一种全新的信息获取和处理技术，在目标跟踪、入侵监测及一些定位相关领域有广泛的应用前景。

WSN 节点定位是根据传感器网络中已经确定了自身位置的已知节点，利用测量得到的未知节点与已知节点之间的相关信息而计算未知节点位置。在 WSN 定位系统中，根据各个节点功能的不同，WSN 节点分为三种类型：中心节点、参考节点以及盲节点。中心节点实现传感器网络与上位机之间的信息交互；参考节点也称为锚节点或信标节点，是 WSN 网络中已知位置信息的节点，提供一定的位置和其他交互信息，协助盲节点实现自定位；盲节点，即 WSN 网络中待定位的节点，该类节点通常是网络中随机部署的节点，其位置可变，可以根据需要随时进行定位。

无线传感器网络定位可以通过不同的方法实现，各种定位方法采用的定位技

术和算法也各不相同，常用的基于距离和角度测量的定位方法有基于信号强度的RSS、TOA、TDOA(Time Difference of Arrival)、AOA等。基于非测距的定位方法主要有质心算法、DV-hop(Distance Vector-hop，距离向量-跳段)算法、APIT(Approximate Point-In-Triangulation Test，近似三角形内点测试)算法等。

4）蓝牙定位技术

蓝牙技术是一种短距离、低功耗的无线传输技术，在室内安装适当的蓝牙局域网接入点，把网络配置成基于多用户的基础网络连接模式，并保证蓝牙局域网接入点始终是这个微微网的主设备，然后通过测量信号强度进行定位，这样就可以获得用户的位置信息。蓝牙技术在移动终端中的应用较为普遍，采用该技术定位较容易发现定位设备，且在大多数非视距的条件下表现良好。但由于其相关设备成本较高且易受噪声干扰，因此系统稳定性较差。

5）A-GPS定位技术

A-GPS(Assisted GPS，辅助GPS)，即辅助GPS技术，是一种结合了网络基站信息和GPS信息对移动台进行定位的技术，可以在GSM/GPRS、WCDMA、CDMA 2000和TD-SCDMA网络中使用。A-GPS需要在手机内部内置GPS模块，并且对手机天线进行相应改造。与传统GPS定位不同的是，手机作为GPS应用设备不需要进行位置信息数据计算，而是将GPS定位数据传输给移动网络，直接由网络定位服务器进行计算。同时移动网络按照GPS的参考网络所产生的辅助数据，如差分校正数据、卫星运行状态等传递给手机，并从数据库中查出手机的近似位置和小区所在的位置信息回传给手机，这样手机就能很快接收到所需要的GPS定位数据，解决了GPS首次定位的时间过长问题，短短几秒钟就能得到所需要的GPS信号和定位信息。

6）UWB定位技术

超宽带技术(Ultra-Wideband，UWB)通过发送和接收具有纳秒或纳秒级以下的极窄脉冲来传输数据，具有GHz量级的带宽。UWB定位技术近几年引起了学术界和业界的极大关注，IEEE 802.15.4a标准将UWB作为定位应用的首选技术。UWB定位通过测距和测向来完成，一般包括三种方法：AOA、RSS、TOA/TDOA等。其中，TOA/TDOA方法是以多径到达时延估计理论为基础的，最能体现出UWB信号时间分辨率高的特点，因此得到了广泛关注和应用。TOA/TDOA方法通过估计到达时延(差)来计算收发两端的距离(差)，充分利用了UWB信号较高的时间分辨率，能体现出UWB高精度定位的优势。超宽带可用于室内精确定位，如室内静止或者移动物体以及人的定位跟踪等方面。目前，由Ubisense公司研发的基于UWB的实时定位系统极具代表性，该系统三维下的精度可达15cm，为目前无线定位最高精度。

1.6 物联网在基层社区矫正信息系统中的应用

社区矫正是目前国际上社会刑罚制度进步的基本趋势。我国实行并发展社区矫正制度，也是与我国的政治、经济、社会及文化相适应的发展要求的表现，对社会和谐稳定，具有非常重要的意义。随着经济社会发展，“社区矫正”逐渐受到各级政府的关注，社区矫正管理系统是利用电子地图技术、通信技术、定位技术，通过移动定位业务(Location Based Service，LBS)服务实现社区矫正人员的位置监控和日常行为管理，实现区域监管、信息交互、警示告知等功能。通过社区矫正信息系统，可以实现对社区矫正对象的信息化、科学化、动态化管理，是推进社会管理创新、强化社区矫正对象管控的一项新举措，将有效提高对社区矫正对象的监管能力和工作效率，防止脱管漏管，推动社区矫正工作迈上新的台阶。

1.6.1 背景介绍

社区矫正(community correction)，也称为“社区矫治”，它是一种不使罪犯与社会隔离并利用社区资源教育改造罪犯的方法，是所有在社区环境中管理教育罪犯方式的总称。

社区矫正本身是一种与传统的监禁刑服刑方式不同的、特殊的服刑方式，法律为“不进监狱”附带了条件。社区矫正的对象必须在判决、裁定宣告的期限内遵守法律及各项规章制度，履行自己的义务，接受有关机构和人员的监督考察，转变犯罪意识和行为恶习。简单地说，就是让符合法定条件的罪犯在社区中执行刑罚。这意味着，只有那些符合社区矫正条件的一部分罪犯，可以在监狱外服刑，而不是所有被判刑的罪犯都不进监狱服刑。

国外较常见的社区矫正包括缓刑、假释、社区服务、暂时释放、学习释放等。我国的“社区矫正”，是指在专门的国家机关，在相关社会团体和民间组织以及社会志愿者的协助下，将符合社区矫正条件的罪犯置于社区内，在判决、裁定或决定确定的期限内，矫正其犯罪心理和行为恶习，并促进其顺利回归社会的非监禁刑罚执行活动。目前，我国正在积极推进社区矫正工作，不断加强对社区矫正对象的教育矫正，通过多种形式，矫正其不良心理和行为，促使其弃恶从善；同时帮助社区矫正对象解决就业、生活、心理及维权等方面的问题和困难。社区矫正工作是国际司法文明潮流的大势所趋，是对中国特色社会主义刑罚制度的探索，彰显了刑罚的社会化、执法的人性化和矫正的文明化，是对传统的监狱内监禁式刑罚方式的重大变革，可以降低刑罚成本，提高刑罚效率和提升罪犯的改造质量，同时也是司法行政工作的一项新职能。

1.6.2　发展动态

1. 国外的社区矫正状况

自 20 世纪以来，在西方法制发达国家，尤其以英美两国成功司法实践经验为模板，广泛兴起了一种新的刑罚执行理念，即社区矫正。可以说，社区矫正这个概念的外延是十分广泛的，它包括一切对于符合社区矫正条件的人员，将其置于社区当中进行开放性教育和帮助，使其矫正行为恶习和犯罪心理的制度设计和方式方法。

以美国为代表的西方国家的社区矫正制度已基本走向成熟，在尝试用最有益的方式处理犯罪和犯罪人的探索中积累了丰富经验。在美国，有两种基本的社区矫正模式。在第一种模式中，综合性的社区矫正方案将量刑规则和司法自由裁量权（俗称“front-end”）与不同种类的替代性刑罚、假释和缓刑考验相结合。在第二种模式中，一些州（包括加利福尼亚）制订了许多方案，根据这些方案，矫治官员将会对已决犯选择适用替代性刑罚、假释和缓刑（又称“back-end”）。这两种模式都被用于帮助缓解监狱过度拥挤的问题，并且相对于监狱服刑来说，廉价的多。

在 20 世纪 70 年代后期，社区矫正作为一种为罪犯提供重返社会的训练所/过渡期宿舍的方法在美国得到了广泛的发展。最初的社区矫正方案在俄勒冈州、科罗拉多州和明尼苏达州作为试验性项目在极少的政府财政拨款支持下开展。在 20 世纪 80 年代后期，针对监狱系统设施严重不足等问题，社区矫正制度在美国 19 个州得到了应用。

2. 我国社区矫正制度现状

我国社区矫正试点始于 2003 年，以 2003 年两高两院印发《关于开展社区矫正试点工作的通知》为界限，可以把我国社区矫正制度的发展划分为两个阶段：第一阶段的特点表现为没有法律依据，刑罚执行机关在司法实践过程中探索适用带有社区矫正性质的制度。自该《通知》公布以来，进入第二阶段，表现为公布了法律依据，正式确立社区矫正的试点工作。2009 年底，全国共有 27 个省（区、市）共有 208 个地（市、州）、1309 个县（区、市）14 202 个乡镇（街道）开展了社区矫正试点工作，累计接收社区服刑人员 35.8 万人，解除矫正 17.1 万人，现有社区服刑人员 18.7 万人；至 2012 年 7 月，社区矫正已在全国 98％的地（市、州）、96％的县（市、区）和 92％的乡镇（街道）开展，全国累计接收社区矫正人员 102 万人，累计解除矫正 56.7 万人，现有社区矫正人员 45.3 万人；截至 2014 年 8 月，全国在册社区服刑人员 70.9 万人，各地累计接收社区服刑人员 184.7 万人，累计解除社区服刑人员 113.8 万人，社区矫正人员再犯罪率一直保持在 0.2％的低水平。同时，司法部研究建立了全国社区矫正人员信息库，各省（区、市）积极探索运用手机定位等电子化监管措

施，加强对社区矫正人员的监督管理。全国社区矫正机构逐渐完善，司法部设立社区矫正管理局，各省（区、市）司法厅（局）设立了社区矫正局（处、办），全国多数地（市、州）和县（市、区）司法局单独设立了社区矫正工作机构。

近年来，我国的社区矫正工作工作从一片空白到试点，再到扩大试点，最后全面实行，取得了较好的效果。2009 年 10 月，两院两部联合下发了《关于在全国试行社区矫正工作的意见》，在全国全面试行社区矫正工作。2010 年全国人大常委会第十九次会议正式审议通过的《刑法修正案（八）》中，社区矫正首次写入刑法。2012 年，为进一步规范社区矫正工作，加强和创新特殊人群管理，最高人民法院、最高人民检察院、公安部、司法部（简称“两院两部”）联合制定了《社区矫正实施办法》，对缓刑、假释以及暂予监外执行罪犯等符合社区矫正条件的人员按照规定程序执行社区矫正，并及时监控和处理各种突发情况，将社区矫正工作落到实处。

社区矫正是一项严肃的刑罚执行活动，其本身就要求必须加强社区规范化建设。十多年来，各地社区矫正工作制度化、规范化建设取得了积极进展。北京、天津、上海、江苏等省（区、市）根据有关法律、法规和“两院两部”《关于开展社区矫正试点工作的通知》、《关于在全国试行社区矫正工作的意见》，制定了社区矫正工作办法或实施细则，对社区矫正工作发展起到了积极作用。与此同时，国内也相继出现了有关社区矫正机构运用信息化技术的范例，如南京社区矫正移动管理信息系统、东营市社区矫正管理信息系统、邢台司法社区矫正系统等，这些系统都是基于移动运营商网络对矫正对象进行管理，基本实现了社区矫正的目的，但是还有很多问题，其中包括各个系统相互独立，没有形成从省部级到镇区级逐级的管理监督，还有开发成本高以及定位精度、定位响应时间、使用范围和终端、系统使用等方面存在不足。

国内社区矫正主要使用的信息化技术包括视频监控、手机汇报、定期访问等。由于各个地区还没有形成健全的、统一的技术规范和技术标准，所以社区矫正工作出现了零散化和信息“孤岛”，不利于今后的信息共享和资源整合。为了进一步规范社区矫正信息化建设，2013 年司法部按照“统筹规划、统一标准、分步实施、分类管理、突出重点、整体推进”的原则，根据《社区矫正实施办法》、国家有关信息化建设规范和相关行业标准，在深入调研论证和广泛征求意见的基础上，制定了《社区矫正管理信息系统技术规范》和《社区矫正人员定位系统技术规范》。这两个规范的主要内容包括社区矫正信息化建设和应用的基本框架、基本流程、数据采集、编码规范、数据交换规范以及系统安全规范等，进一步明确了全国社区矫正信息化建设和应用的总体要求和基本框架，规定了社区矫正管理信息系统的基本功能要求，为研发社区矫正工作相关业务应用系统提供了依据，为实现全国社区矫正信息的资源共享和交换奠定了技术基础。

3. 存在问题

社区矫正是一项全新的工作，目前存在的主要问题包括：

(1) 各项工作主要依赖于文书的记录以及直接信息的传递，这样对于工作的开展造成很大的不便和不安全隐患，同时对数据的处理造成很大的困难和繁重的工作量。

(2) 社区矫正工作执行过程中存在强制力不够、矫正对象的活动范围无法有效监管等问题，导致对矫正对象的监控出现脱管、漏管等现象，严重影响矫正的严肃性和实施效果。

(3) 没有形成标准化、量化考核，由于新的《社区矫正实施办法》刚出台，所以之前的矫正方式都不符合规范，矫正对象的信息管理，以及考核都没有统一的格式和统计规范，造成很难预测矫正效果。

由于矫正对象需要被限制在一定的地域范围内活动，司法行政部门需要掌握矫正对象是否出了限制区域，及时了解矫正对象的大概位置。随着信息化时代的到来，为了更好地促进社区矫正工作向智能化、人性化、效率化，提高司法行政管理的信息化水平，推动社区矫正工作管理的进步，提供了技术基础。根据社区矫正工作的特点，结合国际先进的移动位置服务技术(Location Based Service，LBS)，探索数字化、智能化的社区矫正管理信息系统，具有重要的现实意义。

1.6.3　物联网在社区矫正中的应用

作为基层司法行政工作最重要的工作职责之一，社区矫正工作一直存在“位置查询难、犯人监督难、流程复杂、手续烦琐、部门协调难”等系列问题。物联网技术让这些难题“迎刃而解”。采用物联网和移动信息化技术，司法行政管理部门通过建立一套规范化、系统化的社区矫正管理综合平台，可以有效实现对监外执行的犯人的人性化管理和监督，提升了社区矫正工作的管理效率。利用物联网技术的信息化平台，可以实现司法行政与社区矫正信息的采集、传输、处理、交换，以及对社区矫正对象等的智能化识别、定位、跟踪、监控和管理，相当于在被监管对象周围构筑了一道无形的网。

基于物联网技术的社区矫正系统结构如图 1-32 所示。整个系统分为 4 层，包括感知层、网络层、平台层和应用层，通过分层建设，达到平台能力及应用的可成长、可扩充，创造面向未来的智慧社区矫正系统框架。以云计算为核心，以物联网为触角，打造面向社区矫正信息化应用平台。

随着物联网、云计算等技术的迅速发展，将推动社区矫正信息化的逐步深入。通过建设社区矫正信息化平台，可以使得司法行政部门管理人员及时准确获取社区矫正人员相关动态信息(如心理信息、身体信息、家庭信息、经济收入、社会关系



图 1-32 基于物联网的社区矫正平台结构

等),分析社区矫正全过程中存在的风险隐患点,掌握司法控矫正对象的活动范围,实现矫正对象监管工作的实时远程监控,帮助各地各级司法行政机构降低刑罚成本、提高刑罚效率;同时,通过信息化平台实现对社区矫正人员的位置监控、信息交互、警示告知、考核管理、档案管理、监控中心、矫正管理、呼叫中心、信息互动、统计分析等功能,从而加大对被矫正人员的监管力度,有效防止脱管、漏管,使社区矫正工作合理化、智能化、效率化,为构建社区矫正安全环境,促进社会和谐稳定发挥职能作用。



云计算与大数据技术

随着“大数据”、“云计算”概念的出现，以互联网为主体的现代信息技术呈现出不断迅猛发展的趋势，这为司法行政信息化以及社区矫正信息化建设提供了重要的技术支撑。利用大数据时代的信息技术，由各级司法行政机关牵头，构建社区矫正数据库和信息共享平台，从而有利于社区矫正工作更便捷、更全面、更科学、更高效地开展。

2.1 云计算与大数据概述

2.1.1 云计算技术

1. 概述

由于计算机技术的不断发展和进步，越来越多的 IT 资源以基础设施的形式被提供给人们使用。互联网用户可以通过一个简单的接口，在任意位置、任意时间获取所需的网络、存储、计算等资源。这种新的应用模式和需求，推动了云计算(Cloud Computing)的发展和广泛应用。

云计算是继 20 世纪 80 年代客户端-服务器模式转变之后的又一种巨变。云计算描述了一种基于互联网的新的 IT 服务增加、使用和交付模式，通常涉及通过互联网来提供动态易扩展而且经常是虚拟化的资源。

2. 概念

云计算概念是由 Google 提出的，对于云计算的概念和内涵，业界还没有一个

共识。

Google 定义云计算是一种基于互联网的，可以按需为计算机、其他设备以及大众用户提供软硬件资源、信息和服务的计算方式。

IBM 认为，云计算是一种共享的网络交付信息服务的模式，云服务的使用者看到的只有服务本身，而不用关心相关基础设施的具体实现。其核心原则是：硬件和软件都是资源并被封装为服务，用户可以通过互联网按需地访问和使用。

美国国家标准与技术研究院（National Institute of Standards and Technology，NIST）定义云计算是一种能够提供可用的、便捷的、按使用量付费的、按需访问网络的模式，能够快速提供可配置的计算机资源（包括网络、服务器、存储、应用软件、服务等），而实现这些功能，则只需要投入少量的管理工作。

维基百科（Wikipedia. com）认为，云计算是一种能够将动态伸缩的虚拟化资源通过互联网以服务的方式提供给用户的计算模式，用户不需要知道如何管理那些支持云计算的基础设施。

市场研究机构 IDC 定义，云计算是一种新兴的 IT 技术发展、部署及发布模式，能通过互联网实时提供产品、服务和解决方案。

总结上述概念可以看出，从狭义的角度来看，云计算是指 IT 基础设施的交付和使用模式，指通过网络以按需、易扩展的方式获得所需的资源（硬件、平台、软件）。提供资源的网络被称为“云”，“云”中的资源在使用者看来是可以无限扩展的，并且可以随时获取，按需使用，随时扩展，按使用付费。这种特性经常被称为像水电一样使用 IT 基础设施。从广义的角度来看，云计算是指服务的交付使用模式，指通过网络以按需、易扩展的方式获得所需的服务。这种服务可以是 IT 和软件、互联网相关的，也可以是任意其他的服务，它具有超大规模、虚拟化、可靠安全等独特功能。

从现有的云计算平台来看，它与传统的单机和网络应用模式相比，具有如下特点：

1）超大规模

“云”都具有相当的规模，谷歌云计算已经拥有 100 多万台服务器，亚马逊、IBM、微软和雅虎等公司的“云”均拥有几十万台服务器。“云”能赋予用户前所未有的计算能力。

2）虚拟化

这是云计算的核心技术之一，包括资源虚拟化和应用虚拟化。每一个应用部署的环境与物理平台没有关系，通过虚拟平台实现对应用的扩展、迁移、备份等，操作通过虚拟化层次完成。

3）通用性

云计算不针对特定的应用，在“云”的支撑下可以构造出千变万化的应用，同一

片“云”可以同时支撑不同的应用运行。

4）按需部署

云计算平台可以按照用户的需求分配资源和计算能力。

5）高可靠性

“云”使用了数据多副本容错、计算节点同构可互换等措施来保障服务的高可靠性，使用云计算比使用本地计算机可靠性提高。

6）高可扩展性

“云”的规模可以动态伸缩，可以实时将服务器加入到现有的服务器机群中，增加“云”的计算能力，满足应用和用户规模增长的需要。

7）高性价比

云计算采用虚拟资源池的方法管理所有资源，物理资源可以使用性能较低的PC组成云，也可以使用大型主机。

综上所述，云计算是分布式计算、互联网技术、大规模资源管理等技术的融合与发展，涵盖了数据中心管理、资源虚拟化、大数据处理、信息安全等重要问题。

3. 特征

结合云计算的应用背景及概念，云计算的特征可以归纳如下：

（1）硬件和软件都是资源，通过网络以服务的方式提供给用户。在云计算中，资源从传统的处理器、存储、网络带宽等物理范畴，扩展到了软件平台、Web服务和应用程序等方面。

（2）资源可以根据需要进行动态扩展和配置。借助虚拟资源池提供弹性服务，可以在极短的时间内为企业提供大量的虚拟服务器的资源，并在任务完成后快速地回收这些资源。

（3）资源在物理上以分布共享方式存在，为云中的用户所共享，但最终在逻辑上以单一整体的形式呈现。

（4）用户按需使用云中的资源，按实际使用量付费。即付即用的方式已广泛应用于存储和网络带宽中，虚拟程度的不同导致了计算能力的差异。

总之，在云计算中软、硬件资源以分布共享的形式存在，可以被动态地扩展和配置，最终以服务的形式提供给用户。用户按需使用云中的资源，只需按实际使用量付费。这些特征决定了云计算区别于自给自足的传统IT运用模式，必将引领信息产业发展的新浪潮。

4. 分类

典型的云计算架构主要分为三个基本层次：基础设施（Infrastructures）层、平台（Platform）层和应用（Application）层，如图2-1所示。

按照云计算平台基本架构和SOA理念，采用垂直到整合、物理到虚拟、独用到

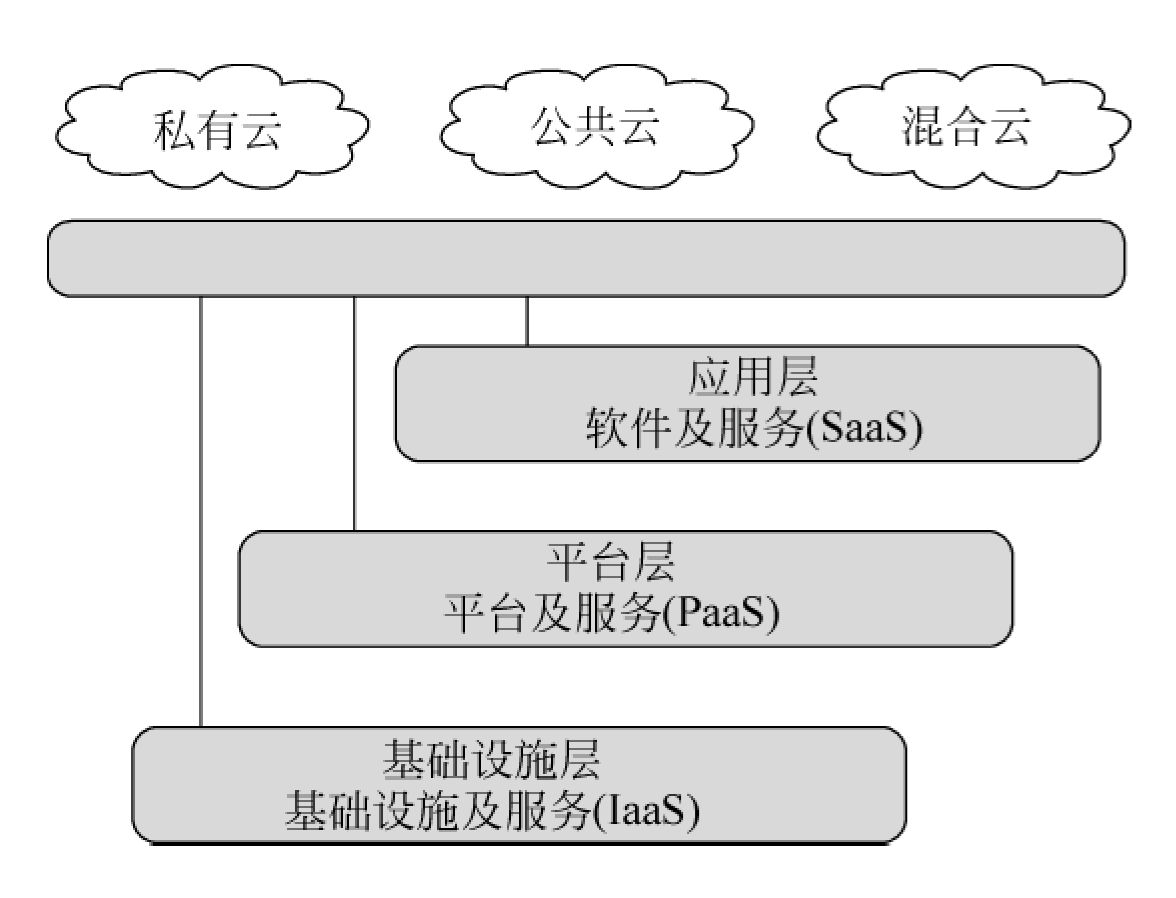


图 2-1 云计算分类

共享、固定到弹性的云计算架构，具体内容包括三层，分别是基础设施云服务(Infrastructure as a Service，IaaS)、平台云服务(Platform as aService，PaaS)和软件云服务(Software as a Service，SaaS)。基础设施云服务的重点是建立全网可管、可控、可调度、多渠道融合的基础设施架构。平台云服务在基础设施云服务平台的基础上，形成全面、规范的系统部署和开发环境。软件云服务重点建设信息云服务门户网、共建共享资源管理系统、云终端服务平台、信息安全监控管理平台等。

1）按服务类型分类

所谓云计算的服务类型，就是指其为用户提供什么样的服务；通过这样的服务，用户可以获得什么样的资源；以及用户该如何去使用这样的服务。目前业界普遍认为，以服务类型为指标，云计算可以分为以下三类。

(1) 基础设施即服务，即 IaaS。IaaS 提供硬件基础设施部署服务，将虚拟硬件资源封装成服务供用户使用，为用户按需提供实体或虚拟的计算、存储和网络等资源。在使用 IaaS 层服务的过程中，用户需要向 IaaS 层服务提供商提供基础设施的配置信息，运行于基础设施的程序代码以及相关的用户数据。IaaS 的最大优势在于允许用户动态申请或释放节点，按使用量计费。由于 IaaS 是供公众共享的，因而资源使用率会较高。为了优化硬件资源的分配，IaaS 层引入了虚拟化技术，借助于 Xen、VMware、KVM 等虚拟化工具，可以提供可靠性高、可定制性强、规模可扩展的 IaaS 层服务。

(2) 平台即服务，即 PaaS。PaaS 对资源的抽象层次更进了一步，它将开发环境作为服务提供给用户，提供应用程序部署与管理服务。通过 PaaS 层的软件工具和开发语言，应用程序开发者只需上传程序代码和数据即可使用服务，而不必关注底层的网络、存储、操作系统的管理问题。

(3) 软件即服务，即 SaaS。SaaS 是基于云计算基础平台所开发的应用程序，

其将具有某些特定功能的应用软件封装成服务，然后通过浏览器将软件传递给成千上万的用户。从用户的角度来说，这意味着他们无须购买软件，只需租用软件；从供应商的角度来说，与常规的软件服务模式相比，他们只需要维护一个软件就行了，这样可以降低成本。对于普通用户来讲，SaaS 层服务将桌面应用程序迁移到互联网，可实现应用程序的泛在访问。

IaaS、PaaS 和 SaaS 三者的比较如表 2-1 所示。

表 2-1　IaaS、PaaS 和 SaaS 的比较

类型	服务内容	服务对象	使用方式	关键技术	系统实例
IaaS	提供基础设施部署服务	需要硬件资源的用户	使用者上传数据、程序代码、环境配置	数据中心管理技术、虚拟化技术等	Amazon EC2\eucalyptus 等
PaaS	提供应用程序部署与管理服务	程序开发者	使用者上传数据、程序代码	海量数据处理技术、资源管理与调度技术等	Google App Engine、Micr-osoft Azure、Hadoop 等
SaaS	提供基于互联网的应用程序服务	企业和需要软件应用的用户	使用者上传数据	Web 服务技术、互联网应用开发技术等	Google Apps、Salesforce CRM 等

2）按服务的方式分类

按照云计算提供者与使用者的所属关系，可以将云计算分为 3 类，即公有云、私有云和混合云。

（1）公有云，是由若干企业和用户共享使用的云环境。在公有云中，用户所需的服务由一个独立的、第三方云提供商提供。该云提供商也同时为其他用户服务，这些用户共享这个云提供商所拥有的资源。

（2）私有云，是由某个企业独立构建和使用的云环境。私有云是指为企业或组织所专有的云计算环境。在私有云中，用户是这个企业或组织的内部成员，这些成员共享着该云计算环境所提供的所有资源，公司或组织以外的用户无法访问这个云计算环境提供的服务。

（3）混合云，指公有云与私有云的混合。

一般来说，对安全性、可靠性及 IT 可监控性要求高的公司或组织，如金融机构、政府机关、大型企业等机构，一般采用私有云方式。这些用户一般都已经拥有了规模庞大的 IT 基础设施，只需少量投资，即可升级自己的 IT 系统，拥有云计算带来的灵活与高效，同时可以有效避免使用公有云可能带来的负面影响。除此之外，也可以选择混合云，将一些对安全性和可靠性需求相对较低的应用部署在公有云上，来减轻对自身 IT 基础设施的负担。相关分析指出，一般中小型企业和创业

公司会选择公有云，而金融机构、政府机关和大型企业则更倾向于选择私有云或混合云。

2.1.2 大数据技术

1. 概述

随着移动互联网、物联网、云计算等技术和应用的兴起，全球范围内数据量迅猛增长，大数据(Big Data)时代已经来临。2011 年 6 月麦肯锡全球研究院发布题为《大数据：下一个创新、竞争和生产力的前沿》研究报告，提出“大数据时代已经到来”。

伴随着大数据的产生，其在现代社会和经济活动中发挥着极其重要的作用，同时有效的利用会使大数据产生不可估量的价值。例如，如果能够充分共享、融合挖掘与分析海量的电子政务数据、移动终端数据、来自物联网传感器的流式数据、企业长期积累的业务数据等，将会产生巨大的经济和社会效益。

大数据是大小超出传统软硬件采集、储存、管理和分析等能力的所有数据集合，它不仅指代“数字”，还统称一切保存在电脑中的信息，包括文本、声音、视频等。大数据的实质是在数据传输、收集、存储的基础上，对数据的分析挖掘，并由此获得凭直觉难以发现的有用信息，从而揭示数据隐藏的历史规律和未来的发展趋势，为决策提供参考。

大数据有四大特点：

(1) 数据量大(Volume)，数据的起始计量单位至少是 P(1000 个 T)、E(100 万个 T)或 Z(10 亿个 T)。

(2) 速度快时高效(Velocity)，这是大数据区分于传统数据挖掘最显著的特征。

(3) 类型繁多(Variety)，包括网络日志、音频、视频、图片、地理位置信息等，多类型的数据对数据的处理能力提出了更高的要求。数据来自多种数据源，数据种类和格式日渐丰富，已冲破了以前所限定的结构化数据范畴，囊括了半结构化和非结构化数据。

(4) 价值密度低(Value)，如随着物联网的广泛应用，信息感知无处不在，信息海量，但价值密度较低，如何通过强大的机器算法更迅速地完成数据的价值“提纯”，是大数据时代亟待解决的难题。

大数据是作为云计算、物联网之后 IT 行业又一大颠覆性的技术革命。云计算主要为数据资产提供了存储、访问的场所和渠道，而数据才是真正有价值的资产。企业内部的生产经营信息、物流信息，互联网中社交信息、位置信息等，其数量将远远超越现有企业 IT 架构和基础设施的承载能力，实时性要求也将大大超越现有的

计算能力。如何盘活这些数据资产，使其为国家治理、企业决策乃至个人生活服务，是大数据的核心议题，也是云计算内在的灵魂和必然的升级方向。

2. 大数据处理技术

大数据的出现对传统的数据存储、数据处理及数据挖掘提出了新的挑战，同时也深刻地影响着人类的生活、工作及思维。传统的数据存储方法、关系数据库、数据处理和数据分析方法已不能满足当前的需要。在大数据处理流程中，最核心的部分就是对于数据信息的分析处理，所以其中所运用到的处理技术也就至关重要。提起大数据的处理技术，就不得不提起“云计算”，这是大数据处理的基础，也是大数据分析的支撑技术。分布式文件系统为整个大数据提供了底层的数据存储支撑架构；为了方便数据管理，在分布式文件系统的基础上建立分布式数据库，提高数据访问速度；在一个开源的数据实现平台上，利用各种大数据分析技术，可以对不同种类、不同需求的数据进行分析整理得出有益信息，最终利用各种可视化技术形象地显示给数据用户，满足用户的各种需求。云计算是大数据分析处理技术的核心原理，也是大数据分析应用的基础平台。

综上，与大数据有关的关键技术有分布式存储与查询、分布式文件系统和分布式计算。

1）分布式存储与查询

典型的分布式存储与查询技术有 BigTable、HBase、Hive。其中 BigTable 由 Google 公司设计，是一种用来处理海量数据的非关系型数据库；HBase 是一个高可靠性、高性能、面向列、可伸缩的分布式数据库，该技术借鉴了 Google BigTable 的思想，Pig 和 Hive 为 HBase 提供了很好的高层语言支持，使得在 HBase 上进行数据统计处理变得非常简单。

2）分布式文件系统

典型的分布式文件系统有 GFS（Google File System）和 HDFS（Hadoop Distributed File System）。GFS 是 Google 开发的一个可扩展的分布式文件系统，用于大型的、分布式的、对大量数据进行访问的应用，运行于廉价的普通硬件上，可提供容错功能，也可给大量的用户提供总体性能较高的服务。HDFS 是开源分布式系统 Hadoop 的一部分，具有高容错性，可部署在低廉的硬件上。数据分块存储在不同的机器上，多台机器拥有同一数据块，以防止某台机器故障导致数据缺失。

3）分布式计算

大数据所涉及的分布式计算大多都是使用 MapReduce 模型，MapReduce 模型的思想来源于函数式编程语言和矢量变量语言，简单易用，可以使编程人员在不了解分布式并行编程的情况下将自己的程序运行在分布式系统中。MapReduce 模型主要包含两步：map 和 reduce。map 主要将一组键值对转换成一组新的中间键

值对，然后 MapReduce 框架会将 map 函数产生的中间键值对里键相同的值传递给一个 reduce 函数；而 reduce 则是接受一个键，以及相关的一组值，将这组值进行合并产生一组规模更小的值(通常只有一个或零个值)。

2.2 云计算平台

2.2.1 云计算平台体系结构

云计算的核心是按需部署，即实现资源的动态可重构、监控和自动化部署等。因此，云计算平台需要具备以下两个特征：

(1) 系统具有自动化的能力，以减轻人工部署和管理负担；

(2) 系统架构应具有快速响应性，能对突发需求做出快速反应。

云计算平台是一个强大的“云”网络，连接了大量并发的网络计算和服务，可利用虚拟化技术扩展每一个服务器的能力，将各自的资源通过云计算平台结合起来，提供超级计算和存储能力。一个典型的云计算平台体系结构如图 2-2 所示。该平台体系结构由用户访问层、服务层和管理层三大部分组成，其中服务层又可细分为资源层、平台层和应用层。

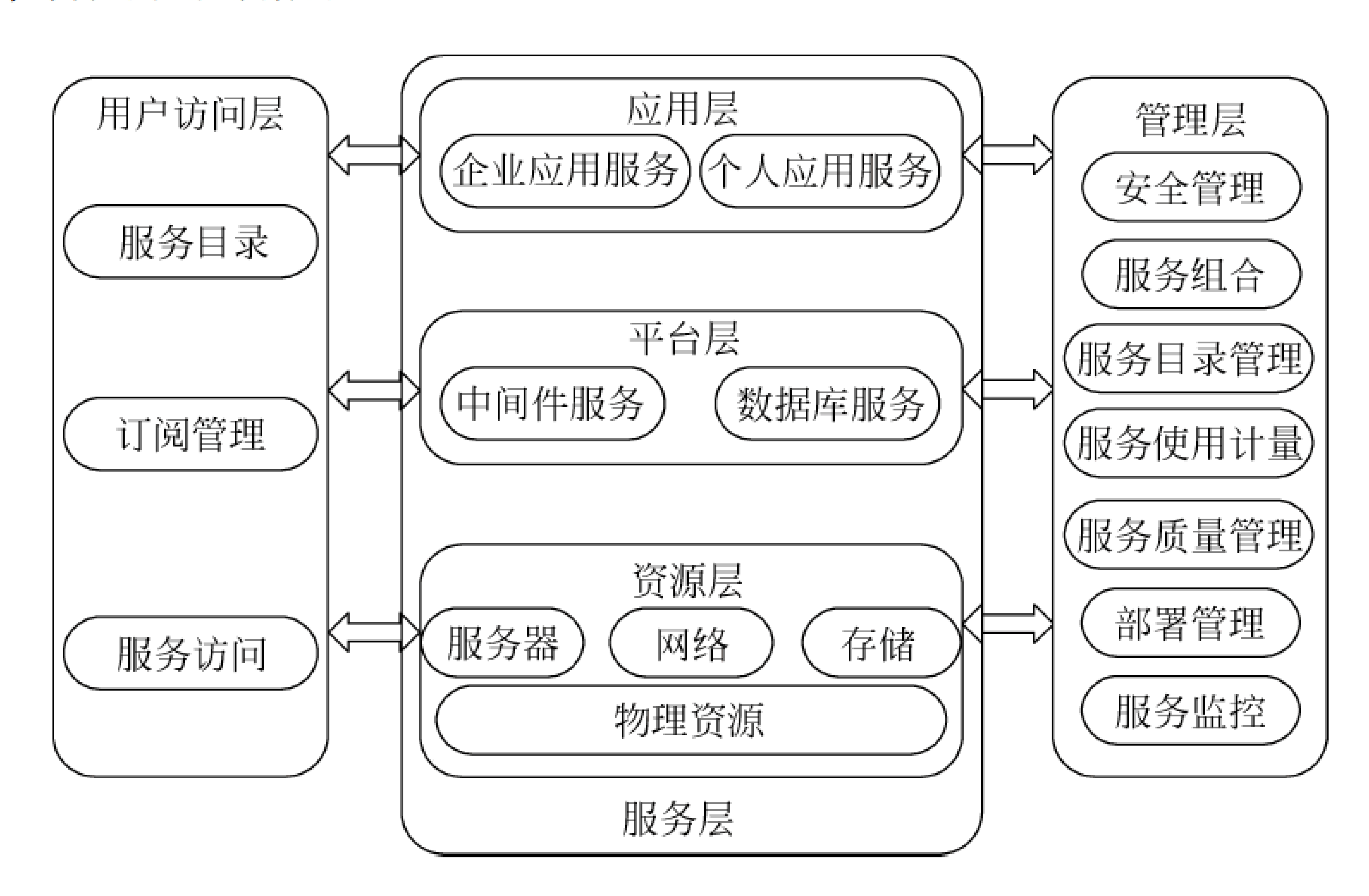


图 2-2 云计算体系结构

各层功能如下：

(1) 用户访问层，这一层是“云”用户请求服务的交互界面，提供给用户访问云计算服务时所需要的接口以及用户使用云计算服务时所需要的必要支撑服务。例如，在服务目录中，用户可以根据自己的需求选择相应的云计算服务；在订阅管理中，用户可以查询自己已订制的服务，或者关闭某个已订制的服务；服务访问就是

针对不同层次的云计算服务为用户提供对应的访问接口。

用户访问接口实现了云计算服务的泛在访问，通常包括命令行、Web 服务、Web 门户等多种形式。其中，通过 Web 门户，云计算将用户的桌面应用迁移到互联网，从而使用户可以随时随地通过浏览器访问数据和程序，以提高工作效率。

(2) 管理层，针对所有层次的云计算服务提供统一的管理，用于管理可用计算资源和服务，能对用户的授权、认证和登录进行管理，接收用户发送的请求，并根据用户请求转发到相应的应用程序。

管理层对应用层的可用性、可靠性和安全性提供保障。云计算需要提供高可靠、高可用、低成本的个性化服务。云计算服务提供商提供的云服务质量(QoS)通过与用户协商后，以服务水平约定(Service Level Agreement，SLA)的形式确定并提供。

此外，数据的安全性一直是用户较为关心的问题，由于云计算数据中心采用资源集中式管理，数据的安全性一直是用户较为关心的问题，由于云计算数据中心采用资源集中式管理的方式，因此存在单点失效、突发事件(如地震、断电)、病毒入侵、非法攻击而丢失或泄露等问题。研究云计算环境下的安全与隐私保护技术(如数据隔离、隐私保护、访问控制等)是保证云计算得以广泛应用的关键。

除了 QoS 保证、安全管理外，服务管理层还包括计费管理、资源监控等管理内容，这些管理措施对云计算的稳定运行同样起到重要作用。

(3) 服务层，将硬件基础设施、软件运行环境、应用程序抽象成服务，这些服务具有可靠性强、可用性高、规模可伸缩等特点，满足多样化的应用需求。具体包括三个层次：资源层、平台层和应用层，其中资源层是将底层物理资源(服务器、存储、网络和负载等)进行池化之后，作为服务提供给用户使用；平台层对资源层服务进行了封装，为用户提供一个应用开发和部署平台，用户可以在该平台之上构建自己的应用；应用层是将应用以 Web 的形式提供给用户。

2.2.2 云计算关键技术

云计算的关键技术主要包括虚拟化技术、数据管理技术、数据存储技术、编程模式及云安全技术。

1. 虚拟化技术

虚拟化技术是云计算实现的关键技术，也是云计算的核心特征。通过虚拟化技术可以将一台或多台物理服务器资源进行池化，然后根据用户业务需求，快速、灵活地进行资源部署。自从 1998 年 VMware 将只有在大型机中采用的虚拟化技术引入 x86 平台至今，服务器虚拟化技术已经得到了广泛应用，新一代的数据中心已经在采用虚拟化技术进行构建。

虚拟化技术主要分为两个层面：物理资源池化和资源池管理。其中，物理资源池化是把物理设备（包括服务器、存储、网络、安全等）由大化小，将物理设备虚拟为多个性能可配的最小资源单位；资源池管理是对虚拟化后所有资源进行管理，根据资源的实际使用情况和用户对资源的申请情况，按照一定的策略对资源进行灵活分配和调度，实现按需分配资源。

在云计算系统中，云计算将每一个层次的功能模块化并且封装成为抽象实体，构建一个动态数据中心，实现软件应用与底层硬件相隔离。其中，当前在云计算平台中应用最为广泛的虚拟化技术主要有 Xen 虚拟机技术以及 KVM（Kernel-based Virtual Machine，基于内核的虚拟机）虚拟机技术。Xen 虚拟机技术是一个基于开源软件组织的虚拟机监控器（Virtual Machine Monitor，VMM），可以允许在单一的物理机器上同时运行多个操作系统实例。KVM 虚拟机技术是一种用于 Linux 内核中的虚拟化基础设施，KVM 目前支持 Intel VT 及 AMD-V 的原生虚拟技术，是基于硬件的完全虚拟化。不同于 Xen，KVM 虚拟化使用 Linux 内核作为它的虚拟机管理程序。

2. 数据管理与存储技术

在云计算系统中，首先，需要对大数据进行计算、存储、读取后进行大量的分析以及在大数据中找到特定的数据等，向用户提供高效的服务。因此，数据管理技术需要能够高效地管理大数据集。其次，云计算系统中的数据管理一般采用列存储的数据管理模式，保证海量数据的存储和分析性能。典型的技术是 Google 的 BigTable 和 Hadoop 团队开发的开源数据管理模块 HBase。BigTable 利用了 Google 文件系统 GFS 所提供的分布式数据存储系统，HBase 利用 Hadoop HDFS（Hadoop distributed filesystem）作为其文件存储系统和利用 Hadoop MapReduce 来处理 HBase 中的海量数据，能够支持数据密集型分布式。

大部分 IT 厂商，包括 Yahoo、Intel 的“云”计划采用的是 HDFS 数据存储技术。未来的发展将集中在超大规模数据存储、数据加密和安全性保证，以及继续提高 I/O 速率等方面。

3. 编程模式

目前，云计算数据的分析与处理普遍采用的是类似 MapReduce 的分布式处理开发框架，MapReduce 通过简单的编程接口为并行处理可划分的大数据提供了支持，向程序员屏蔽了任务调度、数据存储和传输等细节，非常适合解决大数据处理问题的伸缩性需求。

MapReduce 是 Google 提出的一个软件架构，用于大规模数据集的并行运算，它由称为 map 和 reduce 两部分用户程序组成，然后利用框架在计算机集群上面根据需求运行多个程序实例来处理各个子任务，然后再对结果进行归并。

4. 云安全技术

云计算虽然改变了服务方式，但是仍然采用了传统的安全模式，涉及身份认证、基础设施保护、信息安全三个方面。与传统的安全技术相比，在云计算平台中，安全设备及手段部署的位置发生了变化，同时服务的安全性由服务提供商来负责保障。

(1) 用户身份安全方面，采用资源授权和认证方式，确保得到授权的用户访问某一应用或系统。

(2) 基础设施安全保护方面，主要是针对软硬件设备(包括网络设施、操作系统、应用系统等)所采用的安全保障措施，确保虚拟机不受非法攻击和入侵。

(3) 信息安全方面，需要确保数据的保密性、完整性。

2.3　大数据处理技术

大数据时代的到来对数据的存储、处理及分析提出了新的挑战，但总的发展趋势是通过分布式计算来解决“瓶颈”问题。具体实现中，Google、Amazon、微软和VMware等公司分别推出了自己的大数据方案，此外还有开源的Hadoop平台。Hadoop是谷歌大数据平台的开源实现，由于其开源特性，越来越多的企业在Hadoop的基础上对其进行修改以适应自己的需要，如Facebook根据其业务需求，底层采用Hadoop平台进行数据的存储和处理，并在其上开发了Hive系统。

2.3.1　大数据处理基本流程

大数据处理系统不管结构如何复杂，采用的技术千差万别，但是总体上可以分为以下几个重要部分。大数据处理基本流程如图2-3所示。

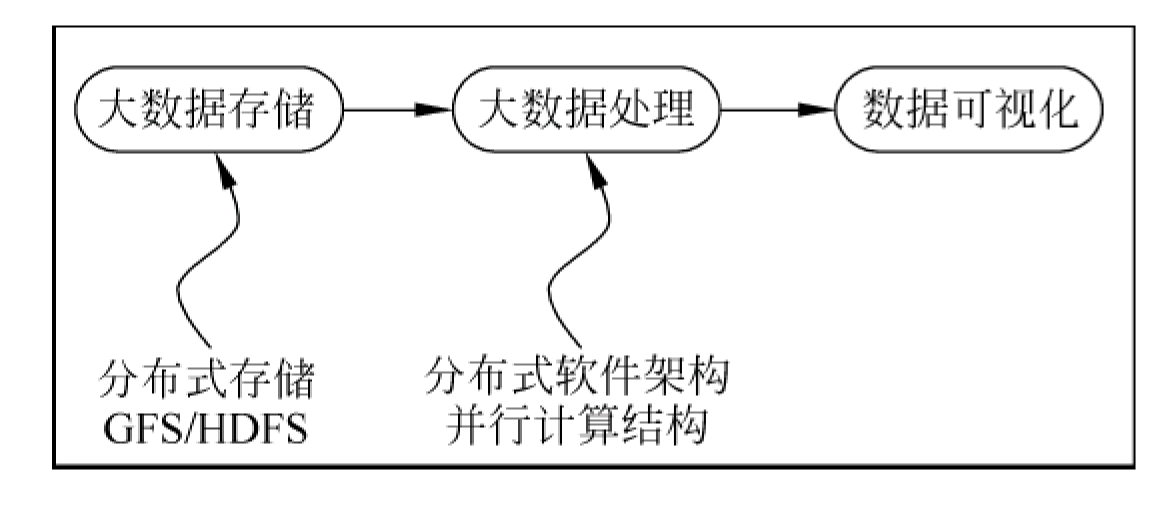


图2-3　大数据系统结构

下面将从支持大数据系统所需要的分布式文件系统、分布式数据处理技术、分布式数据库系统和开源的大数据系统Hadoop等方面介绍大数据系统的关键技术。

2.3.2 大数据关键技术

1. 分布式文件系统

文件系统是支持大数据应用的基础，典型的分布式文件系统有 GFS（Google File System）和 HDFS（Hadoop Distributed File System）。

1）GFS

GFS 是 Google 开发的一个可扩展的分布式文件系统，用于大型的、分布式的、对大量数据进行访问的应用，运行于廉价的普通硬件上，可提供容错功能，也可给大量的用户提供总体性能较高的服务。GFS 主要采用主从（Master-Slave）结构，通过数据分块、追加更新等方式实现海量数据的高速存储。GFS 体系结构如图 2-4 所示。

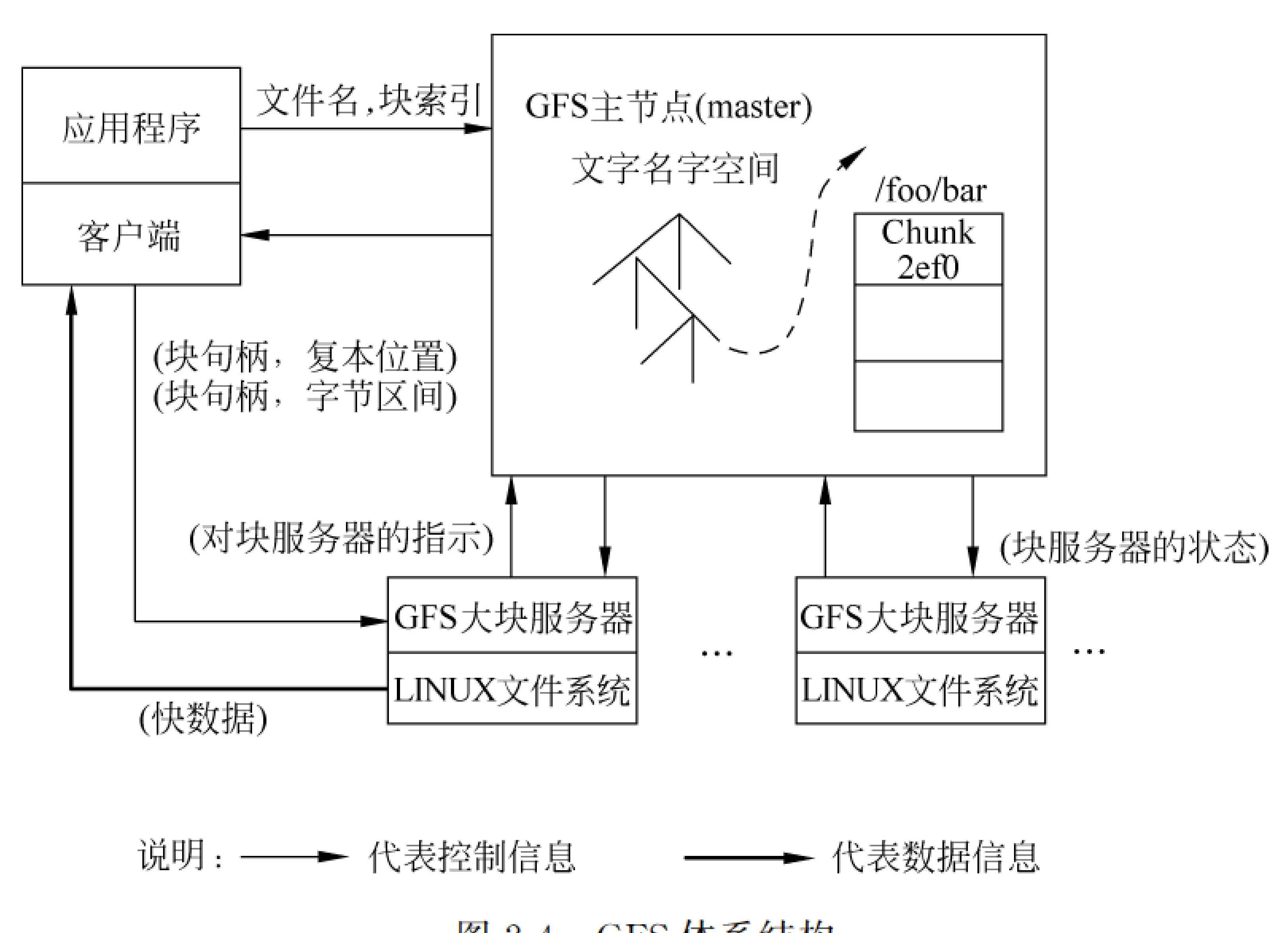


图 2-4 GFS 体系结构

一个 GFS 集群由一个总控服务器 GFS Master 和大量的数据块服务器 Chunk Server 构成，并被许多 GFS 客户（Client）访问。GFS 文件被分成固定大小的块（chunk），每个块由一个不变的、全局唯一的 64 位的 chunk-handle 标识，chunk-handle 是在块创建时由 master 分配的。块以普通的 Linux 文件形式存储在本地磁盘中，并可以读和写由 chunk-handle 和位区间指定的数据。为了保证可靠性，每一个块被复制到多个 chunkserver 上，默认为三份。

Master 中维护了系统的元数据，包括文件及块名字空间、文件到块之间的映射、块位置信息等，也负责整个系统的全局控制，如块租约管理、垃圾回收无用块、块复制等。Master 会定期与 CS 通过心跳的方式交换信息。

Client 是 GFS 提供给应用程序的访问接口，它是一组专用接口，不遵守 POSIX 规范，以库文件的形式提供。Client 访问 GFS 时，首先访问 Master 节点，获取与之进行交互的块服务器信息，然后直接访问这些块服务器，完成数据存取工作。GFS 将整个系统的节点分为三种角色：GFS Master（总控服务器），GFS Chunkserver（数据块服务器，简称 CS）以及 GFS Client（客户端）。

GFS 文件被划分为固定大小的数据块（Chunk），由 Master 在创建时分配一个 64 位全局唯一的 Chunk 句柄。CS 以普通的 Linux 文件的形式将 Chunk 存储在磁盘中。为了保证可靠性，Chunk 在不同的机器中复制多份，默认为三份。

Master 中维护了系统的元数据，包括文件及 Chunk 名字空间，GFS 文件到 Chunk 之间的映射，Chunk 位置信息。它也负责整个系统的全局控制，如 Chunk 租约管理、垃圾回收无用 Chunk、Chunk 复制，等等。Master 会定期与 CS 通过心跳的方式交换信息。

Client 是 GFS 提供给应用程序的访问接口，它是一组专用接口，不遵守 POSIX 规范，以库文件的形式提供。Client 访问 GFS 时，首先访问 Master 节点，获取与之进行交互的 CS 信息，然后直接访问这些 CS，完成数据存取工作。

2）HDFS

HDFS（Hadoop Distributed File System）是开源分布式系统 Hadoop 的一部分，为 Hadoop 提供高性能、高可靠、高可扩展的存储服务。数据分块存储在不同的机器上，多台机器拥有同一数据块，以防止某台机器故障导致数据缺失。

HDFS 体系结构如图 2-5 所示。

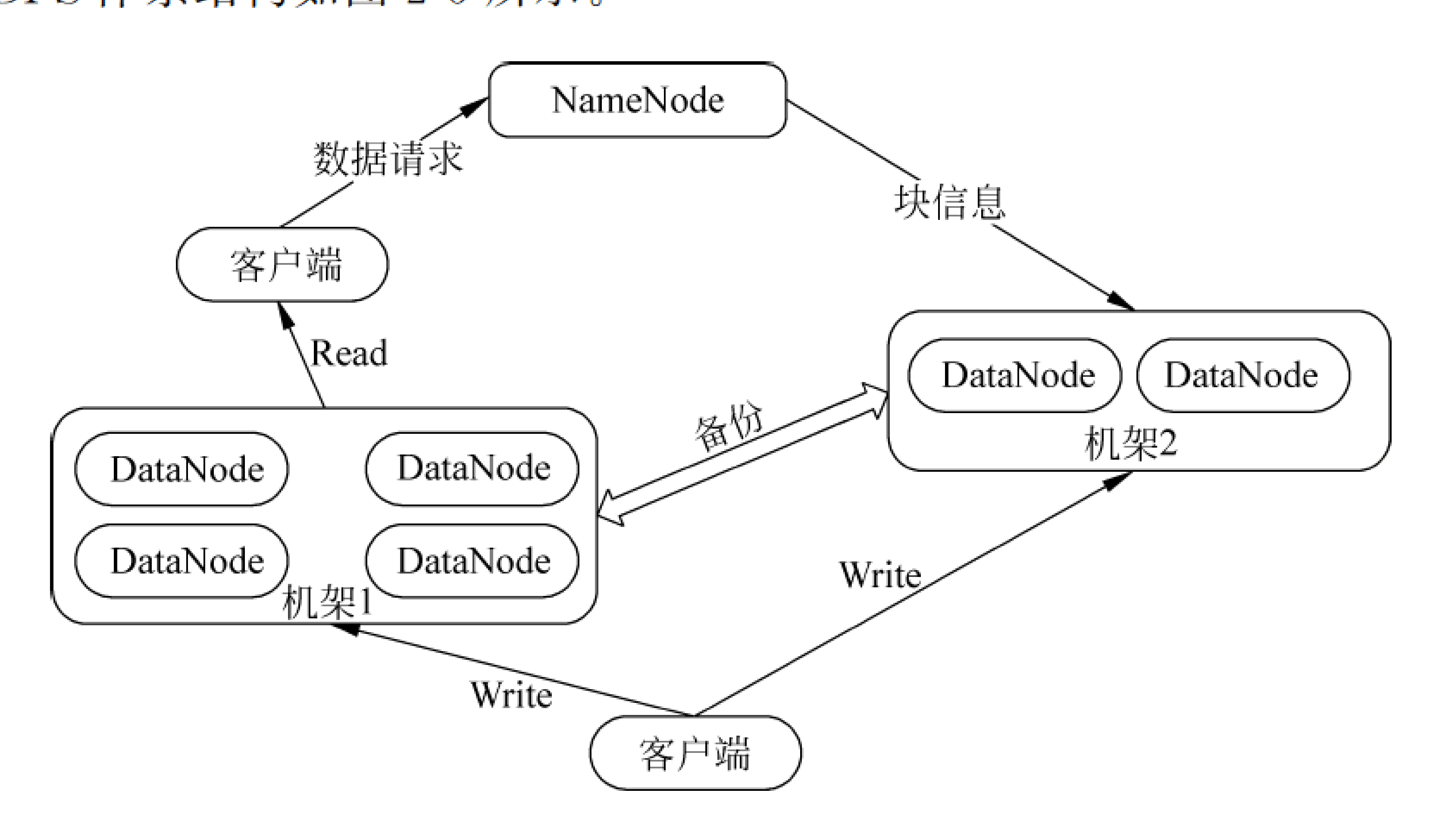


图 2-5　HDFS 体系结构

HDFS 采用了主从结构，包括一个 NameNode（主节点）和多个 DataNode（从节点），并提供应用程序访问接口。NameNode 作为主服务器，管理文件系统的命

名空间和客户端对文件的访问操作。DataNode 提供真实文件数据的存储服务。HDFS 有两种数据形式，分别是文件数据和元数据。HDFS 将用户文件分成若干个数据块，并存放在一组 DataNode 上。NameNode 执行文件系统的命名空间操作，比如打开、关闭、重命名文件或目录等，它也负责数据块到具体 DataNode 的映射。DataNode 负责处理文件系统客户端的文件读写请求，并在 NameNode 的统一调度下进行数据块的创建、删除和复制工作。HDFS 还提供一个分级的文件组织形式，维护这个文件系统所需的信息(除了文件的真实内容)称为 HDFS 的元数据。元数据由 NameNode 进行维护和管理，NameNode 在启动时，会从磁盘加载元数据到内存，并且等待 Data Node 上报其他的元数据信息，形成最终的元数据结构。由于 NameNode 是单节点，一旦 NameNode 无法正常服务，将导致整个 HDFS 无法正常服务。

2. 分布式数据处理系统

大数据的处理模式分为流处理和批处理两种。流处理是直接处理，批处理采用先存储再处理。流处理将数据视为流，源源不断的数据形成数据流。当新的数据到来即立即处理并返回所需的结果。目前比较有代表性的开源流处理系统主要有 Twitter 的 Storm、Yahoo 的 S4 以及 Linkedin 的 Kafka 等。Google 公司于 2004 年提出的 MapReduce 编程模型是最具代表性的批处理模型。MapReduce 模型的主要贡献就是通过简单的接口来实现自动的并行化和大规模的分布式计算，通过使用 MapReduce 模型接口实现在大量普通的 PC 上的高性能计算。MapReduce 对于大数据处理的基本构思是分而治之，将大数据任务分解为多个子任务，将得到的各个子结果组合并成为最终结果。

MapReduce 对大数据的处理可抽象为两个主要阶段：在 Map 阶段先对初始的键-值(Key/Value)对进行处理，产生一系列的中间结果 Key/Value 对；然后再通过 Reduce 阶段合并所有具有相同 Key 值的 Key/Value 对，得到最终结果。

MapReduce 对数据进行处理的应用思路如图 2-6 所示。

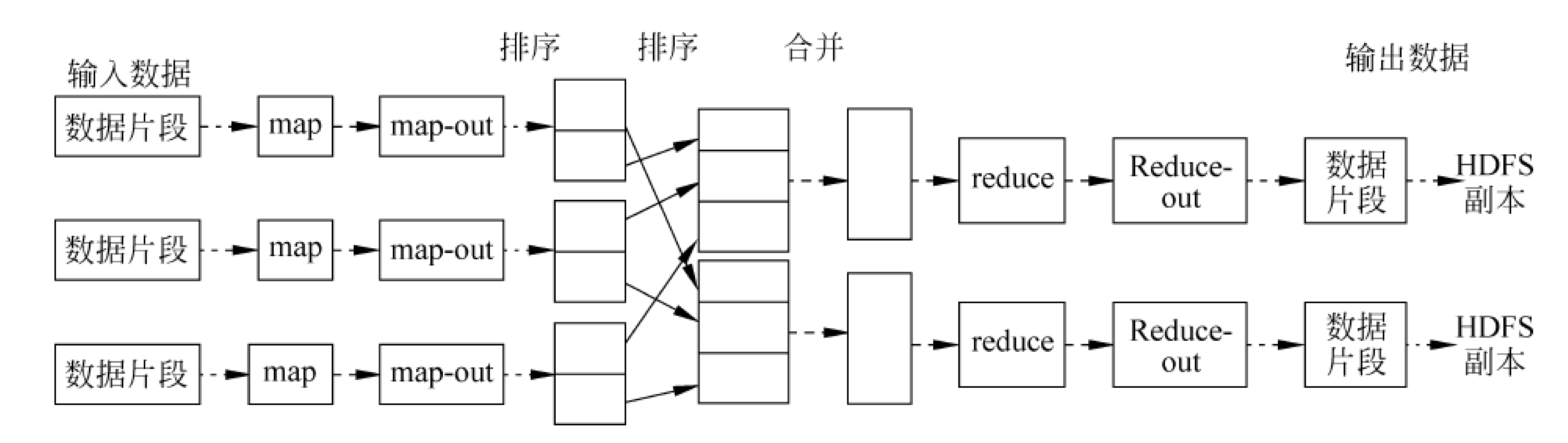


图 2-6　MapReduce 对数据进行处理的应用思路

MapReduce 并行处理流程(待处理的大数据被分为大小相同的块)主要步骤如下:

(1) 用户作业程序提交给主节点;

(2) 主节点为作业程序寻找和配备可用的 Map 节点和 Reduce 节点;

(3) 主节点启动 Map 节点执行程序,读取本地数据;

(4) 每个 Map 节点处理读取的数据块,将中间结果放在本地并通知主节点计算完成及结果数据存储位置;

(5) 主节点启动 Reduce 节点运行,远程读取中间结果并处理。

MapReduce 在系统层面解决了大数据分析平台的扩展性和容错性问题,是非关系型数据库的典型代表,因此越来越多的研究人员从性能和易用性方面对 MapReduce 进行改进。对 MapReduce 性能提升的研究包括 4 个方面:

(1) 多核硬件与图形处理器上的性能改进;

(2) 索引技术与连接技术的优化;

(3) 调度技术优化;

(4) 其他优化技术。

针对 MapReduce 易用性的研究成果包括 Yahoo 的 Pig、Microsoft 的 LINQ、Hive 等。

3. 分布式数据库系统

传统的关系模型分布式数据库难以适应大数据时代的要求,原因主要在于:大数据时代的数据远远超出单机处理能力,分布式技术是必然的选择;大数据时代数据类型的多样性和低价值密度性,在大数据时代数据的存在的形式是多样的,各种半结构化、非结构化的数据是大数据的重要组成部分;设计理念的冲突,关系数据库追求的是“一种尺寸适用于所有”,但在大数据时代不同的应用领域在数据理性、数据处理方式以及数据处理时间的要求上千差万别,不可能存在一种统一的数据存储方式适应所有场景。

面对这些挑战,以 Google 公司推出的 BigTable 为代表未采用关系模型的 NoSQL(Not only SQL)数据库由此诞生,NoSQL 数据库具有模式自由、备份简易、接口简单和支持海量数据等特性,对于实现大数据的存储和处理十分有效。

Bigtable 的设计目的是可靠的处理拍字节(Petabyte,PB)级别的数据,并且能够部署到千台机器上。Bigtable 已经实现了以下几个目标:适用性广泛、可扩展、高性能和高可靠性。Bigtable 不支持完整的关系数据模型,为用户提供了简单的数据模型,利用这个模型,客户可以动态控制数据的分布和格式。BigTable 主要是一个分布式多维表,表中数据通过行关键字、列关键字和时间戳来进行索引和查询定位,并且 BigTable 将存储在表中的数据视为字串,具体数据结构的实现由用户

自行定义。

BigTable 的基本构架如图 2-7 所示，BigTable 中的数据以子表形式保存在子表服务器上，最终以 GFS 文件形式存储在文件系统中。客户端程序直接和子表服务器通信，Chubby 服务器完成对子表服务器的状态监控，主服务器通过查看 Chubby 服务器目录来终止出现故障的子服务器并将其数据转移至其他子服务器。另外，主服务器还完成子表的创建和负载均衡等操作。

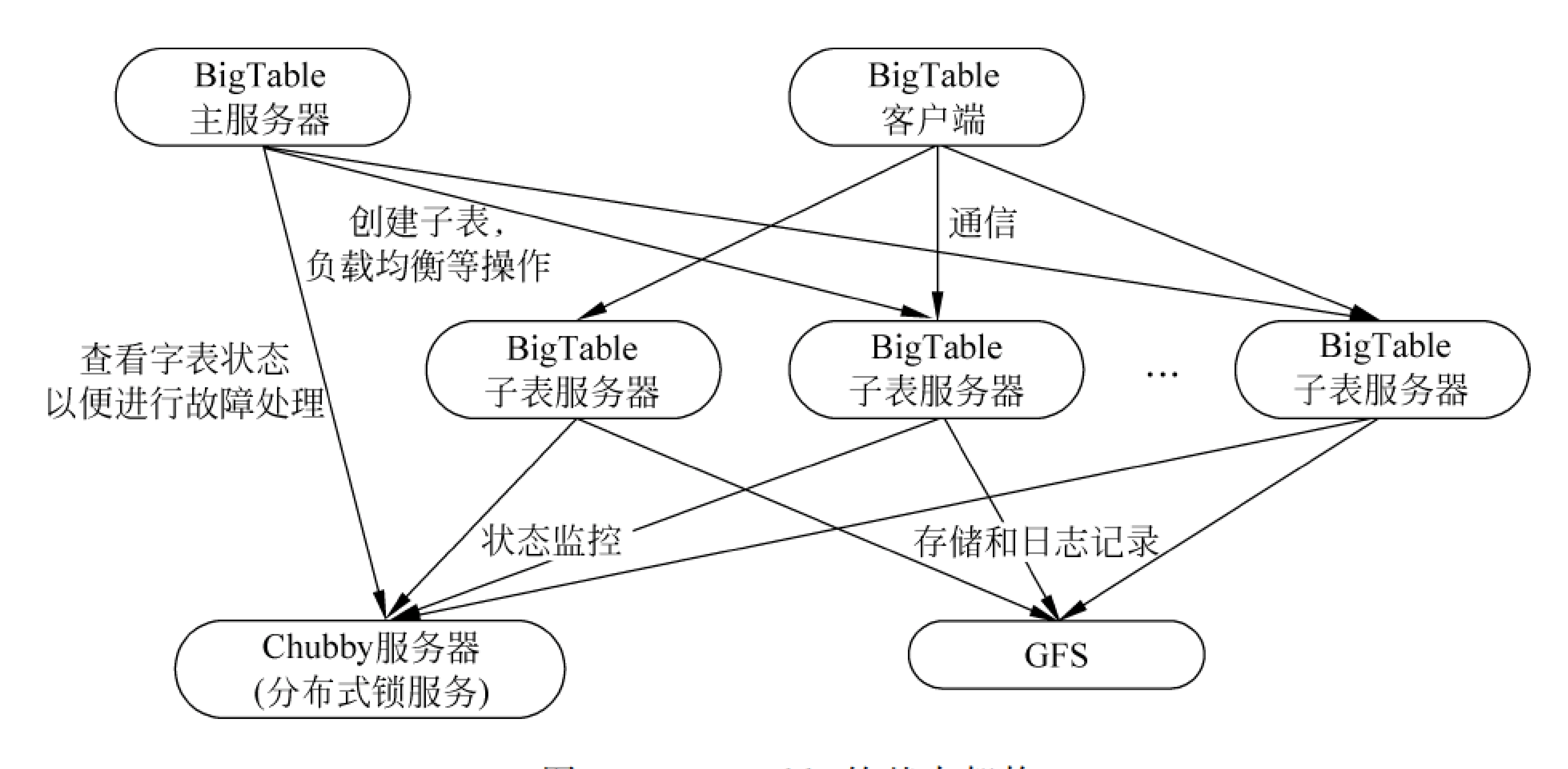


图 2-7 BigTable 的基本架构

除了 Google 公司的 Bigtable，具有代表性的系统有 Amazon 的 Dynamo 和 Yahoo 的 PNUTS。Dynamo 综合使用了键/值存储、改进的分布式哈希表(DHT)、向量时钟等技术实现了一个完全的分布式、去中性化的高可用系统。PNUTS 是一个分布式的数据库系统，在设计上使用弱一致性来达到高可用性的目标，主要的服务对象是相对较小的记录，比如在线的大量单个记录或者小范围记录集合的读和写访问，不适合存储大文件、流媒体。Bigtable、Dynamo、PNUTS 等技术的成功促使研究人员开始对关系数据库进行反思，产生了一批为采用关系模型的数据库，这些方案通称为 NoSQL(not only SQL)。

2.3.3 大数据系统的开源实现平台 Hadoop

Hadoop 起源于 Apache Nutch，2006 年 2 月，NDFS 和 MapReduce 从 Nutch 转移出来，成为一个独立的 Lucene 子项目，称为 Hadoop。Hadoop 是开放源码并行计算编程工具和分布式文件系统，是 MapReduce 的开源实现，凭借其开源和易用的特性，成为大数据处理的首选。此外，还提供构建在 HDFS 和 MapReduee 之上的可扩展的数据仓库 Hive、结构化数据库 HBase、数据流高层语言 Pig、高性能分布式协同服务 ZooKeeper 以及面向大规模分布式系统的数据收集软件 Chukwa

等。到目前为止,Hadoop 技术已经在互联网领域得以广泛应用,同时也得到研究界的普遍关注。

Hadoop 是一个实现了 MapReduce 计算模型的开源分布式并行编程框架,以 Hadoop 分布式文件系统 HDFS 和 MapReduce 为核心的 Hadoop 为用户提供了系统底层细节透明的分布式基础架构。HDFS 的高容错性、高伸缩性等优点允许用户将 Hadoop 部署在低廉的硬件上,形成分布式系统;MapReduce 分布式编程模型允许用户在不了解分布式系统底层细节的情况下开发并行应用程序。所以用户可以借助 Hadoop 框架及云计算核心技术 MapReduce 来实现数据的计算和存储,融合 HDFS 文件系统和 HBase 数据库,实现云计算的分布式、并行计算和存储,并且得以实现很好的处理大规模数据的能力。Hadoop 框架如图 2-8 所示。

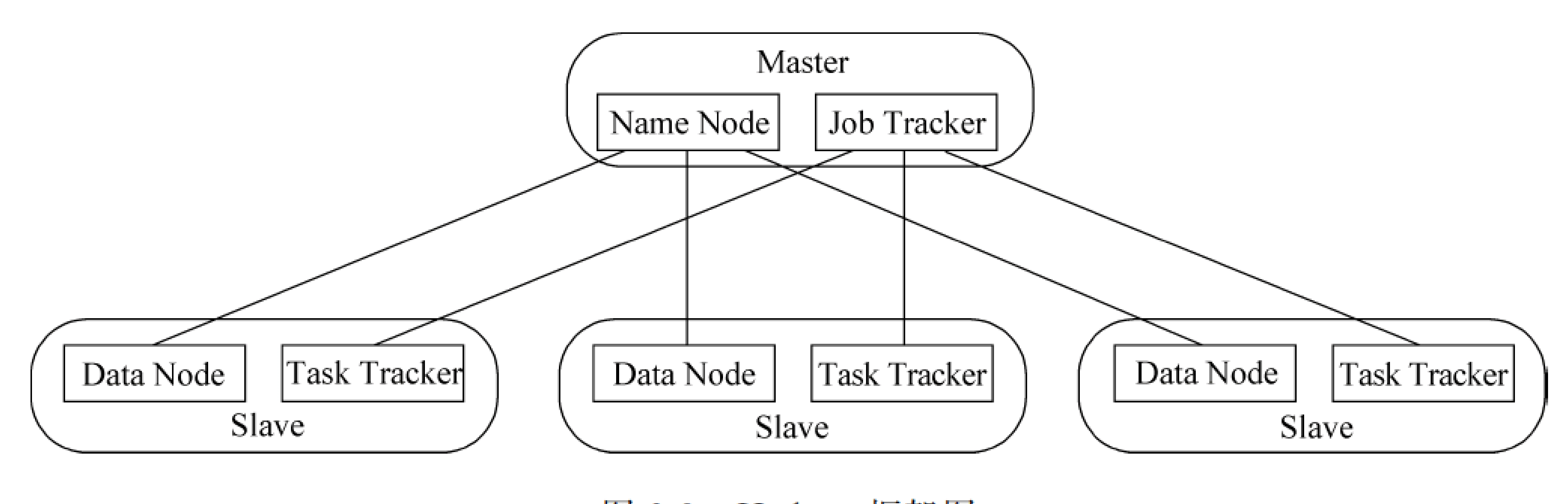


图 2-8　Hadoop 框架图

如图 2-8 所示,在 Hadoop 的系统中,有一台 Master,主要负责 Name Node 的工作以及 Job Tracker 的工作。Job Tracker 的主要职责就是启动、跟踪和调度各个 Slave 的任务执行。还会有多台 Slave,每一台 Slave 通常具有 Data Node 的功能并负责 Task Tracker 的工作。Task Tracker 根据应用要求来结合本地数据执行 map 任务以及 reduce 任务。

2.4　典型云计算支撑环境与工具

云计算技术近几年受到了 IT 厂商的关注,国内外各大 IT 企业、运营商也纷纷将目光聚焦于此,研发自己的相关产品或开拓自己的相关技术。典型的云计算平台有 IBM 的“蓝云”、Google 的云计算和 Amazon 的 EC2 等平台。

2.4.1　IBM 蓝云

“蓝云(Blue Cloud)”是由 IBM 云计算中心开发的企业级云计算解决方案,可以帮助企业对现有的基础架构进行整合,通过虚拟化技术和自动化技术,构建企业自己的云计算中心,实现企业软硬件资源的统一管理、统一分配、统一部署、统一监

控和统一备份，提高资源利用率，从而帮助企业实现云计算理念。

“蓝云”软件平台的特点主要体现在虚拟机以及对于大规模数据处理软件 Apache Hadoop 的使用上。“蓝云”采用的虚拟化方式有两个级别，一个是实现硬件级别上的虚拟化，另一个是通过开源软件实现虚拟化。硬件级别的虚拟化可以使用 IBM p 系列的服务器，获得硬件的逻辑分区 LPAR(logic partition)。然后，通过 IBM Enterprise Workload Manager 来管理逻辑分区的 CPU 资源，采用这种方式以及在实际使用过程中的资源分配策略，能够使相应的资源合理地分配到各个逻辑分区。p 系列系统的逻辑分区最小粒度是 1/10 颗 CPU。在蓝云计算平台中使用了 Xen 虚拟化软件，Xen 是一个开源的虚拟化软件，能够在 Linux 平台上运行另外一个操作系统。

“蓝云”存储体系结构包含类似于 Google File System 的集群文件系统以及基于块设备方式的存储区域网络 SAN(Storage Area Network，存储区域网络)。在设计云计算平台的存储体系结构时，可以通过组合多个磁盘来获得较大磁盘容量。同时，针对多个磁盘同时进行数据读写以及带来的读写速度问题，有两种存储技术，一个是使用类似于 Google File System 的集群文件系统，另一个是基于块设备的存储区域网络 SAN 系统。在“蓝云”计算平台上，SAN 系统与分布式文件系统(如 Google File System，GFS)并不是相互对立的系统，SAN 提供的是块设备接口，需要在此基础上构建文件系统，才能被上层应用程序所使用。而 Google File System 正好是一个分布式的文件系统，能够建立在 SAN 之上。两者都能提供很好的可靠性、可扩展性。“蓝云”基本架构如图 2-9 所示。

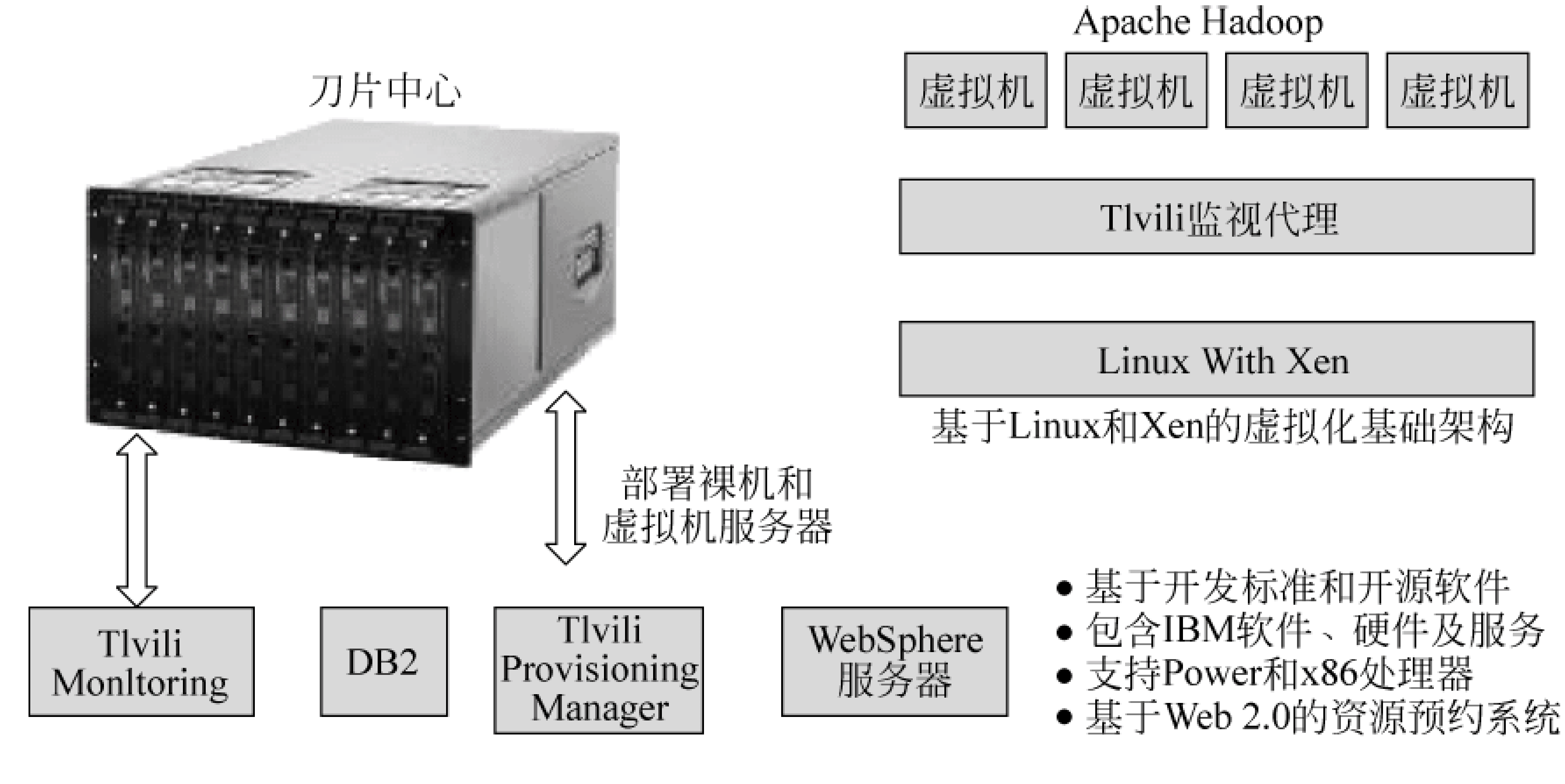


图 2-9 IBM 蓝云的基本架构

2.4.2　Google 的云计算平台

Google App Engine(GAE)是 Google 公司推出的云计算服务,允许用户使用 Python 编程语言编写 Web 应用程序在 Google 的基础架构上运行。另外,Google App Engine 还提供了一组应用程序接口(API),主要包括 datastore API、images API、mail API、memcache API、URL fetch API 和 user API。用户可以在应用程序中使用这些接口来访问 Google 提供的空间、数据库存储、E-mail 和 memcache 等服务,用户可以通过 Google App Engine 提供的管理控制平台管理用户 Web 应用程序。

简言之,Google App Engine 是一个由 Python 应用服务器群、BigTable 数据库及 GFS 组成的平台,它能为开发者提供一体化的、主机服务器及可自动升级的在线应用服务。Google App Engine 专为开发者设计,开发者可以将自己编写的在线应用运行于 Google 云平台上,应用运行时所需要的资源以及应用运行维护由云平台提供。

GAE 的组件调用关系如图 2-10 所示。

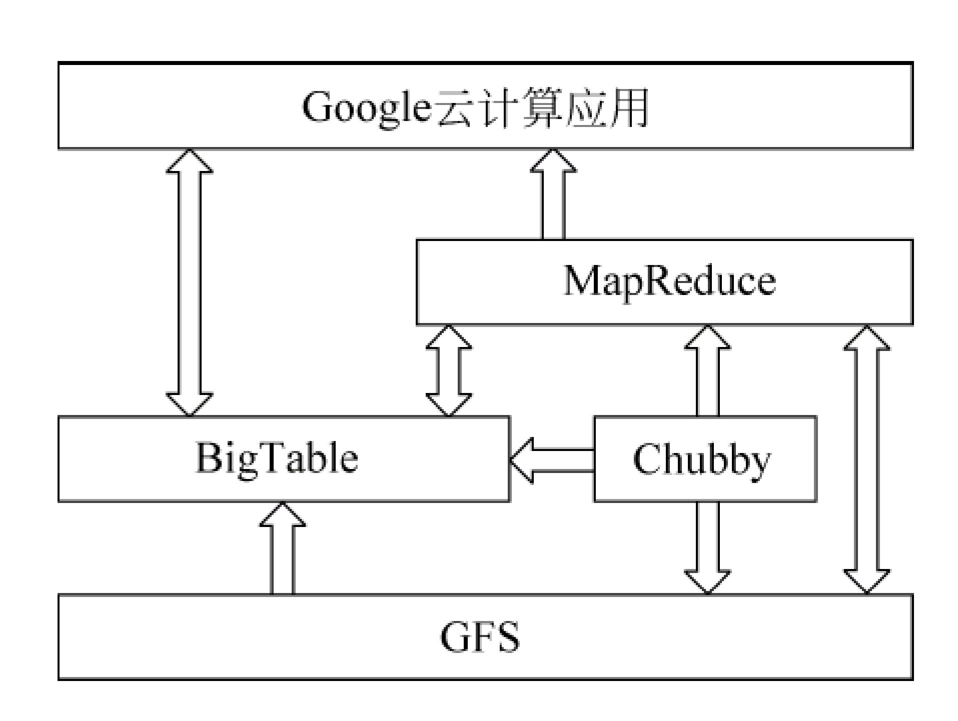


图 2-10　GAE 的组件调用关系

2.4.3　Amazon AWS

Amazon 公司构建的云计算平台以 Web 服务的方式提供云计算服务,包括了从 IaaS 到 SaaS 多个级别的各种云服务,Amazon Web Services(AWS)是这些 Web 服务的总称。通过 AWS 的 IT 基础设施层服务和丰富的平台层服务,用户可以在 Amazon 公司的云计算平台上构建各种企业级应用和个人应用。

AWS 基础设施层服务包括 Elastic Computing Cloud(EC2)、Simple Storage Service(S3)、Simple DB、简单队列服务(Simple Queue Service, SQS)、弹性 MapReduce 服务、内容推送服务 CloudFront、电子商务服务 DevPay、灵活支付服务(Flexible Payment Service,FPS)等。

(1) 弹性计算云 EC2,向用户提供一个运行在 Xen 虚拟化平台上的基于 Linux 的虚拟机,用户可以在此之上运行基于 Linux 的应用程序。EC2 的主要特征有灵活性、低成本、安全性、易用性和容错性。

(2) 简单存储服务 S3,为任意类型的文件提供临时或永久的存储服务。非关系数据库存储模式具有简单、高效等特点。

(3) 简单数据库服务 Simple DB,用以管理复杂的结构化数据,支持数据的查找、删除、插入等操作。SDB 采用树状结构,没有事务的概念,不支持连接操作,实际存储的数据类型比较单一,查询结果只包含条目名称而不包括相应属性值,返回结果不支持排序操作。SDB 增添了一些新特征,如无需预定义模式,单个属性允许有多个值,支持自动索引。

(4) 简单队列服务 SQS,目标是解决低耦合系统间的通信问题,支持分布式计算机系统之间的工作流,其特点是简单、无处不在。

2.4.4 Microsoft Windows Azure

微软公司推出了两种云部署模型,即公有云和私有云。其中,公有云是面向外部用户需求,通过开放网络提供云计算服务。微软公司支持公有云的主要产品是 Windows Azure 云计算平台,该平台具有高可扩展性,提供了即用即付款的灵活服务模式,由一个公共平台上的多种不同服务组成,主要包括微软的云服务操作系统以及一组为开发人员提供的接口服务。Windows Azure 的主要目标是为开发者提供一个平台,帮助开发可运行在云服务器、数据中心、Web 和 PC 上的应用程序。云计算的开发者能使用微软全球数据中心的存储、计算能力和网络基础服务。Azure 平台提供的服务主要有 Live Services、.NET 服务、SQL 数据库服务、SharePoint 服务以及动态 CRM 服务。同时,Azure 平台支持多个 Internet 协议,主要包括 HTTP、REST、SOAP 和 XML,从而为用户提供一个开放、标准以及能够互操作的环境。Windows Azure 平台结构如图 2-11 所示。

在图 2-11 中,最底层的是 Windows Azure 操作系统,它提供了 Compute(计算)、Storage(存储)以及 Manage(管理)三个主要功能。此外,还有对用户而言透明的 Fabric,Fabric 包含负载平衡、硬件抽象等众多功能。在此基础上,提供了中间件产品 AppFabric、数据库产品 SQL Azure 以及其他一些 building block 产品。开发人员可以直接在 Windows Azure 之上进行开发,也可以利用 AppFabric、SQL Azure 等产品的各种特性进行开发。

私有云平台为企业提供了构建企业私有云服务的平台,为企业内部需求提供云计算、数据等服务。基于微软环境的私有云平台主要基于 Windows Server Hyper-V 和 System Center 等软件产品来构建,其逻辑架构如图 2-12 所示。

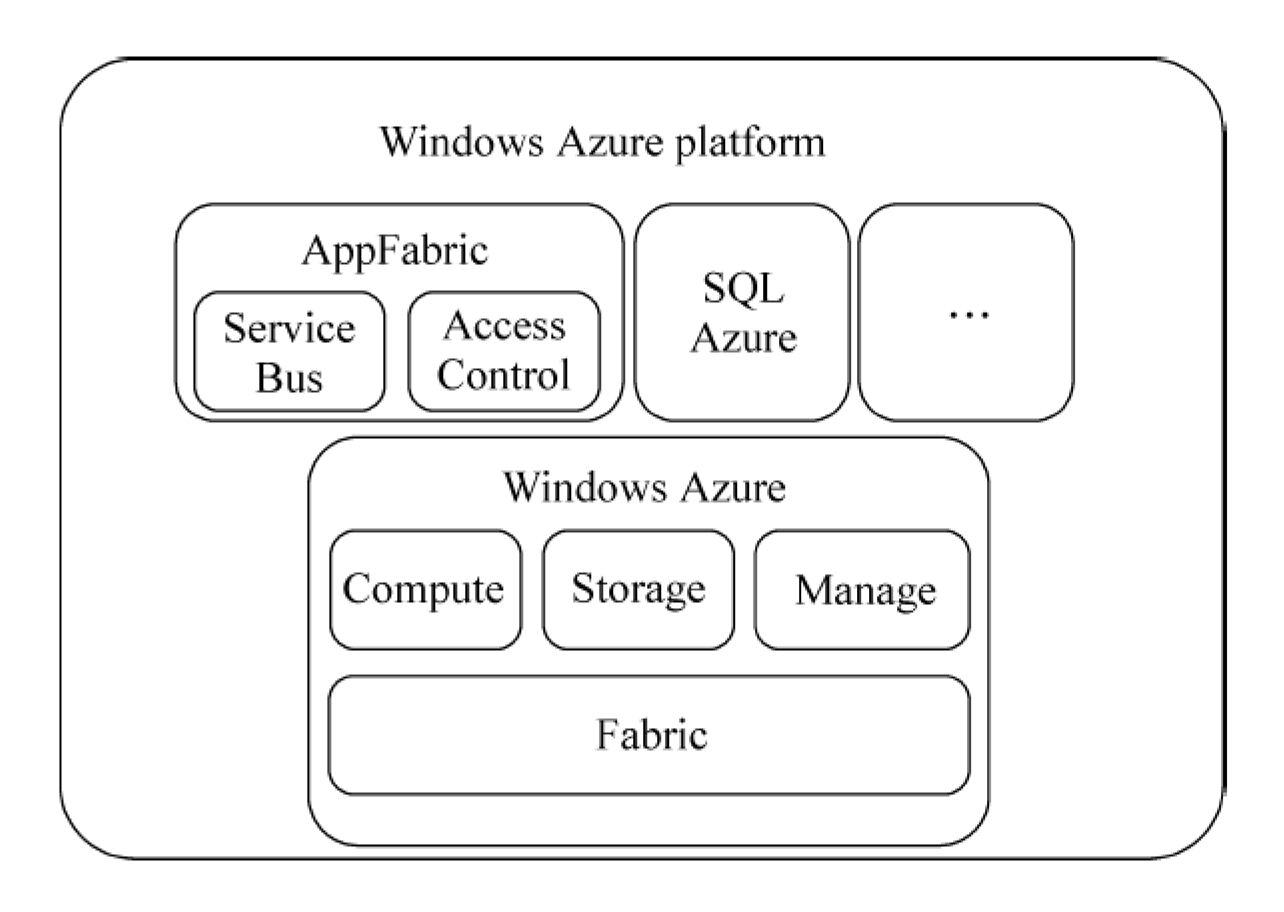


图 2-11　Windows Azure 平台结构

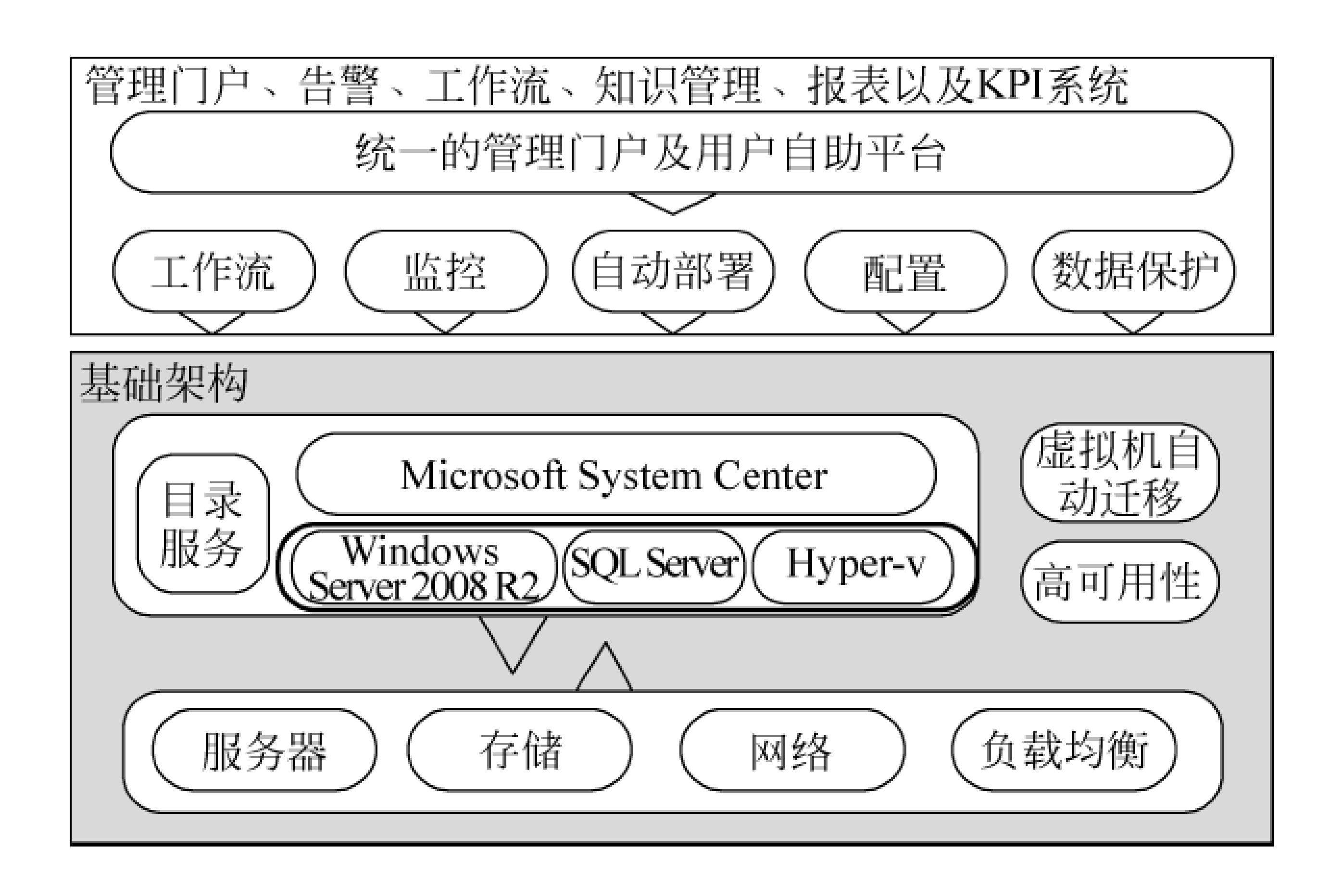


图 2-12　微软私有云平台逻辑架构

如图 2-12 所示的逻辑架构分为四个部分：底层是资源层，之上是使用 Hyper-v 进行资源虚拟化的虚拟层，在虚拟层之上是使用 System Center 对物理环境和虚拟环境进行统一管理的管理层，最上层则是服务层，用户通过该层与管理层和虚拟层建立连接来使用资源层所提供的各种资源。其中虚拟层是整个动态数据中心的基础核心层。

微软私有云解决方案的核心是动态的数据中心架构，通过该架构企业或机构能够快速地部署面向内部使用的私有云平台，服务提供商可基于该架构构建云平台对外提供服务。与此同时，微软在私有云架构中提出了两种解决方案，即面向企业客户的方案(System Center Virtual Machine Manager Self-Service Portal 2.0，SSP 2.0)和面向服务提供商的方案(Dynamic Data Center Toolkit for Hoster，DDTK-H)。这两种解决方案都包括了服务层中所涉及的工作流、监控、自动部署、

配置和数据保护等功能模块。用户可以通过统一的管理门户或用户自助平台接入。

2.4.5 基于 IaaS 的三种开源云平台

作为云计算的一种重要形式,IaaS 服务有上面介绍的几种典型云平台方案外,还有各种开源云平台。下面介绍几种典型的开源 IaaS 云平台。

1. OpenStack

OpenStack 是由 Rackspace 和 NASA 共同开发的云计算平台,可帮助服务商和企业内部实现类似于著名云计算平台 Amazon EC2 和 S3 的云基础架构服务。OpenStack 包含两个主要模块:Nova 和 Swift,前者是 NASA 开发的虚拟服务器部署和业务计算模块;后者是 Rackspack 开发的分布式云存储模块,两者可以一起用,也可以分开单独用。OpenStack 是开源项目,除了有 Rackspace 和 NASA 的大力支持外,还有包括 Dell、Citrix、Cisco、Canonical 等著名公司的贡献和支持,其发展速度非常快。

与其他开源 IaaS 云平台相比,OpenStack 拥有较高的社区开放人气和庞大的生态系统,支持 OpenStack 的企业数量已经达到一百多家。不过,现阶段的产品成熟度存在短板,远远不及 Eucalyptus,同时 OpenStack 的配置和部署相对复杂,其管理和控制界面也不够完善,对使用者自身的技术实力和资源投入有较高的要求。

2. CloudStack

CloudStack 是一个具有高可用性及扩展性的开源云计算平台。目前 CloudStack 支持主流的 hypervisors(虚拟机管理器),如 KVM、XenServer、VMware、Oracle VM、Xen 等。

同时 CloudStack 是一个开源云计算解决方案,可以加速高伸缩性的公共和私有云(IaaS)的部署、管理、配置。以 CloudStack 为基础,数据中心操作者可以快速方便的通过现存基础架构创建云服务。

CloudStack 进入 Apache 阵营之前,在商业领域进行了长期的积累,因此能够为用户提供良好的用户界面和丰富的功能,用户体验较好,同时商业用户部署较为便捷。

3. Eucalyptus

Eucalyptus 源于美国政府支持的加州大学圣巴巴拉分校研究项目,现已转为 Eucalyptus System 公司商业化运作,但是其核心代码仍是开源的。

Eucalyptus 平台的 API 全面兼容 Amazon 的 API,为已经拥有虚拟化环境的用户提供了功能强大的二次开发环境,为服务提供商及企业用户创建私有云或混

合云提供了基础设施服务解决方案。

2.5 云计算应用迁移与部署技术

2.5.1 云迁移的基本原理

1. 什么是云迁移

云迁移是将企业或组织的数字资产、服务、IT 资源以及应用程序等全部或部分从现有数据中心迁移到云环境中的整个过程。

图 2-13 为云迁移技术的原理示意图，下面从迁移对象、迁移手段、迁移目标三个方面来阐述云迁移的基本原理。

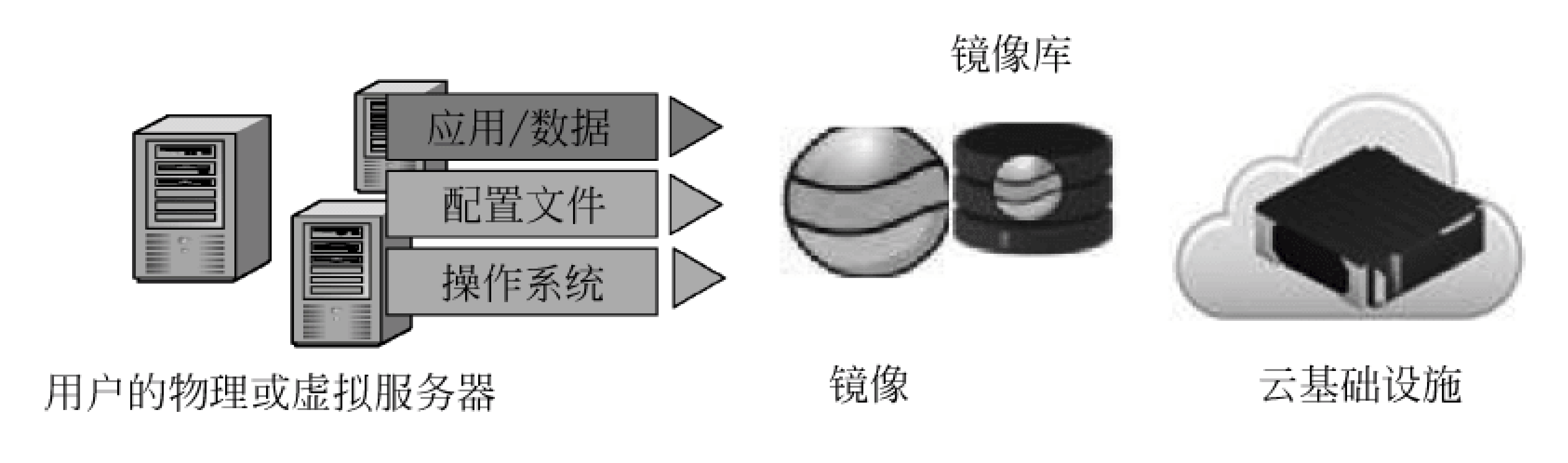


图 2-13　云迁移原理

1）迁移对象

云迁移技术的核心实施对象是现已存在且正在提供某种服务的软件。这些软件部署于企业或团体组织内部，在传统的物理服务器上运行，或在虚拟化的数据中心运行。为了让核心业务软件在新的环境中正常运转，除了业务软件本身之外，迁移的对象还包括该业务软件所依赖的软件运行时的环境。

在迁移对象建模中，如图 2-14 所示，首先需要确定的是迁移的核心业务软件，如大型计算类应用、Web 访问类应用、信息管理系统、特定行业应用软件等；第二层为业务软件 A 正常运行所需的数据，这些数据可能存在于关系型数据库中，也可能以普通文件或特定的文件格式存在；第三层为业务软件 A 运行时所依赖的其他应用软件或中间件服务，他们之间存在着协同操作或服务和被服务的关系，如果缺少了这些支撑软件，业务软件 A 的某些功能就无法运行；第四层为系统服务软件，这些软件为服务器级别的常规管理软件提供最基础的底层支撑服务，如数据库服务器、Web 服务器、安全管理软件如身份认证服务等；最后一层为操作系统，负责管理底层硬件资源，为上层应用提供硬件级的资源使用服务，是应用程序正常运行的环境，所以必须要划分到迁移对象的范畴。

2）迁移手段

迁移手段解决如何运输迁移对象的问题，即在确定好完整有效的迁移对象之

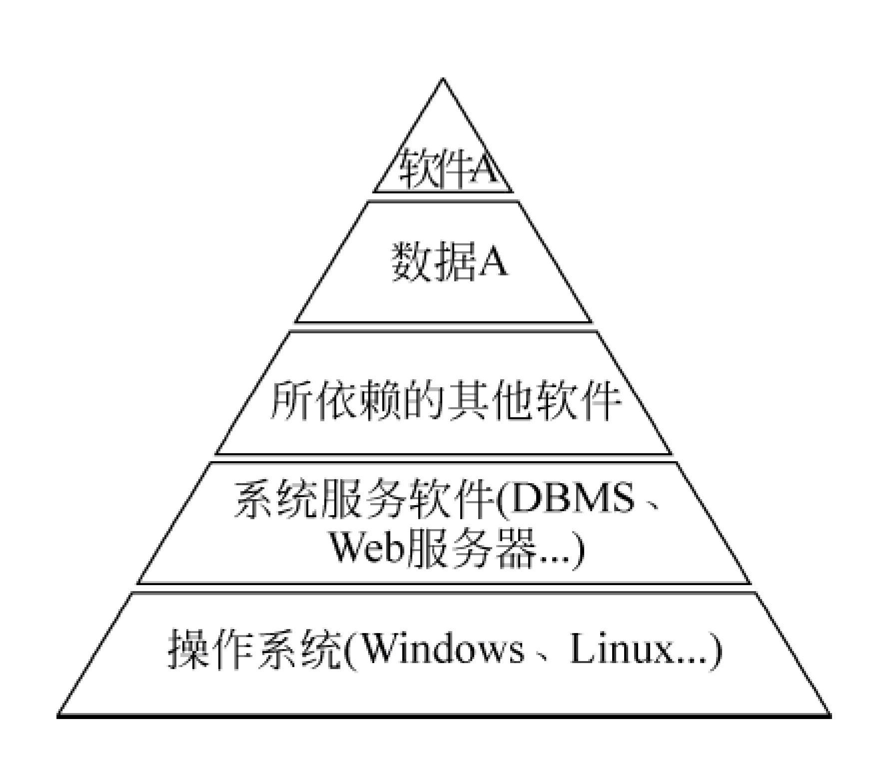


图 2-14　迁移对象模型

后，需要考虑以何种方式将迁移对象从当前的物理环境或虚拟化环境中迁出。当前的主流技术是以镜像（Image）为载体来装载迁移对象。也就是说，云迁移技术致力于把确定好的迁移对象制作成一个机器镜像，以镜像文件的方式来保存和运输迁移对象。不同的迁移对象分别装入不同的机器镜像中，存储在镜像库中，可以作为软件在云环境中运行的启动模板。这种以镜像作为迁移手段的方式，一方面得益于日益流行的云服务模型 IaaS，另一方面依托于虚拟化技术，尤其是服务器虚拟化技术。此外，随着 PaaS 服务和 SaaS 服务的发展，云迁移有了新的实施手段，即将业务托管给 PaaS 服务提供商或者直接使用 SaaS 服务替代原有业务，例如将数据库托管给云数据库。

3）迁移目标

迁移对象被制作成机器镜像后，可随时以镜像为模板，以 IaaS 云服务提供商提供的虚拟化的硬件资源启动虚拟机，启动相关服务，这样业务软件便可在云环境中运行了。相对于原有的传统物理环境的软件运行模式，这种镜像方式的云环境运行模式提高了应用的可移植性。镜像作为软件的原始模板可以随时在不同环境、不同地点使用，而不用受限于初次部署时的环境。但这仅受益于单纯的虚拟化技术，并没有利用真正意义上的云计算服务的特性和优势。云迁移的最终目标是要适当调整应用程序，合理分配云环境中的动态资源，使得在最小开销范围内，保证业务的安全和高可用性。

（1）弹性可扩展，灵活应对业务峰值。

IaaS 级云计算服务将硬件资源虚拟化成虚拟资源池，供用户按需使用。因此在有限的范围内，硬件资源可以灵活地进行弹性伸缩，以满足业务弹性变化的需求。这种弹性扩展包括两种方式：一种是横向的扩展，指虚拟机数量上的扩展。例如初始阶段开启 n 台虚拟机，在业务量达到某一特定阈值时，将虚拟机扩展到 m 台（$m>n$），业务量下降时，又可关掉部分虚拟机；另一种是纵向的扩展，指虚拟机配置上的扩展，数量不变。这种弹性的横向或纵向的扩展特性，大大提高了应用程

序应对业务访问的灵活性。横向弹性扩展意味着应用程序在云环境中的工作模式，将变为集群式的或多虚拟机并行的运行模式，然而大部分现有应用的软件架构都是以单服务器模式运行，在迁移到云环境中后，就需要解决集群模式下各项服务的协同通信、数据的一致性、负载均衡等问题。因此，如何调整现有应用的架构，在最小入侵的情况下，保证应用程序能够适应云环境相对于传统环境的差异性是云迁移技术需要解决的核心问题。

（2）提高资源利用率，降低经济投入。

公有云模式下使用云服务要按需付费。因此云迁移必须要考虑到经济问题，合理地评估业务对虚拟资源的需求，制定云资源的使用方案，应用运行过程中实时监控云资源的使用情况，实时优化资源的使用方案，在确保应用性能的情况下尽可能地缩减服务使用费用。而私有云模式下，则要求要尽可能地提高资源利用率，减少用户的基础设施投入。

（3）安全及尽可能高的性能。

多数企业或用户对云计算服务望而却步的原因主要有两点。一是安全问题。一方面应用程序和相关业务数据从企业的防火墙内部迁移到公共的云平台上，安全控制只能依托于云服务商提供的管理服务；另一方面云计算服务以多租户的模式共享云中的资源，不同租户间的业务运行在同一平台上。这便存在着很大的安全隐患。二是性能问题。应用程序运行在虚拟化环境中的性能要低于传统的物理环境。此外，云环境中不同服务之间的通信通过网络来传递，普通用户也需要通过网络来访问云端服务，网络延迟是影响云端应用性能的主要因素。云迁移技术考虑在现有的云计算技术之上，通过调整应用架构，合理规划云资源的分配，设计迁移策略来尽可能地降低安全风险，规避云应用的性能瓶颈。

总而言之，云迁移技术是监管一切云迁移过程中的活动，研究如何确定云迁移对象，如何运输云迁移对象，如何分配和使用云资源，如何监管和优化云应用的运行等相关的技术和实现。

2. 机器镜像（Image）

镜像（机器镜像）是装载迁移对象的载体。从功能上看，镜像是虚拟机的启动模板，它描述且包含了在虚拟化环境中运行的虚拟机的所有组件。具体而言，镜像是软件模块、系统软件、应用软件以及某个时期的配置信息的一个固定集合。一个机器镜像可以按需部署在虚拟化的环境中，而不用考虑提供给它 CPU、内存和数据存储资源的物理硬件。从格式上看，原始镜像是依赖于云平台的底层虚拟化管理软件（Virtual Machine Monitor），它类似于仅装载了操作系统的磁盘文件。

原始镜像制作完成后，要被压缩、分割、注册最终上传到云平台上供用户使用。云平台对上传镜像进行统一管理，包括镜像的存储方式、命名 ID、相关信息文件，

以及提供统一的访问工具来供用户使用该镜像。

2.5.2 云迁移策略

迁移策略是指迁移的方法模式、技术手段等。相对于传统物理环境，云环境具有它自身的独特之处。其一，云计算的核心技术是虚拟化，IaaS 级服务将硬件（计算、网络、存储）资源抽象成虚拟资源池，以虚拟机的方式提供服务。这意味着所有迁入云中运行的应用都必须要以虚拟化的方式运行，因此首先需要辨别应用是否适合以虚拟化方式运行。其二，云计算主张对资源的支配要按需进行，并能够根据需求的变化而灵活地、弹性地扩展和收缩。

一般来说，云迁移策略秉持两大基本准则：

（1）充分了解云计算服务模型，全面合理地利用云服务资源。

云计算提供三种类型的服务：IaaS、PaaS、SaaS。IaaS 级服务提供虚拟机、网络服务、存储服务这类硬件资源，为应用在云中运行提供坚实的硬件服务，虚拟机也可以直接为应用程序提供运行环境。PaaS 级服务提供平台级服务，如云环境中的开发、部署、监测等工具以及其他云中的平台级软件，来帮助应用程序更方便、快捷、有效地迁移入云。PaaS 级服务为应用程序迁移进云环境后的重构或再制造提供了可能性。SaaS 级服务提供软件级服务，是迁移入云的应用程序获取其他软件服务的来源。因此需要充分了解云计算服务模型，熟悉云服务提供商所提供的各项服务，尽可能全面地考虑利用云服务本身的资源，从而制定出合理有效的云迁移策略。

（2）最小化的入侵应用程序。

应用的入侵性是指对现有的、已经投入使用的应用程序的部署方式、架构设计、程序代码等内容的修改。

大多数现有应用由于其自身架构差异或其他方面的原因，不能直接在云环境中运行。如果放弃原有应用去重新开发适合云环境的新版本，会大大增加用户开发或购买软件的成本，还需要投入大量人力物力，在时间上也无法尽快满足业务需求。理想情况下，现有应用无须修改或调整就能在云环境运行，在云迁移的实施过程中要尽量最小化应用修改和重构，保证原有应用的功能和完整性，同时尽量减少云迁移带来的额外经济投入。

图 2-15 中的云符号表示对应层在云环境中运行。针对不同的迁移对象及云服务模型的不同层次，可把云迁移策略归纳为以下四种方式。

1. 整体迁移

整体迁移是指将应用程序的整个软件栈迁移到云环境中。这种方式是云迁移的经典方式，也是当前普遍流行的一种方式。整体迁移立足于当前已经相对成熟

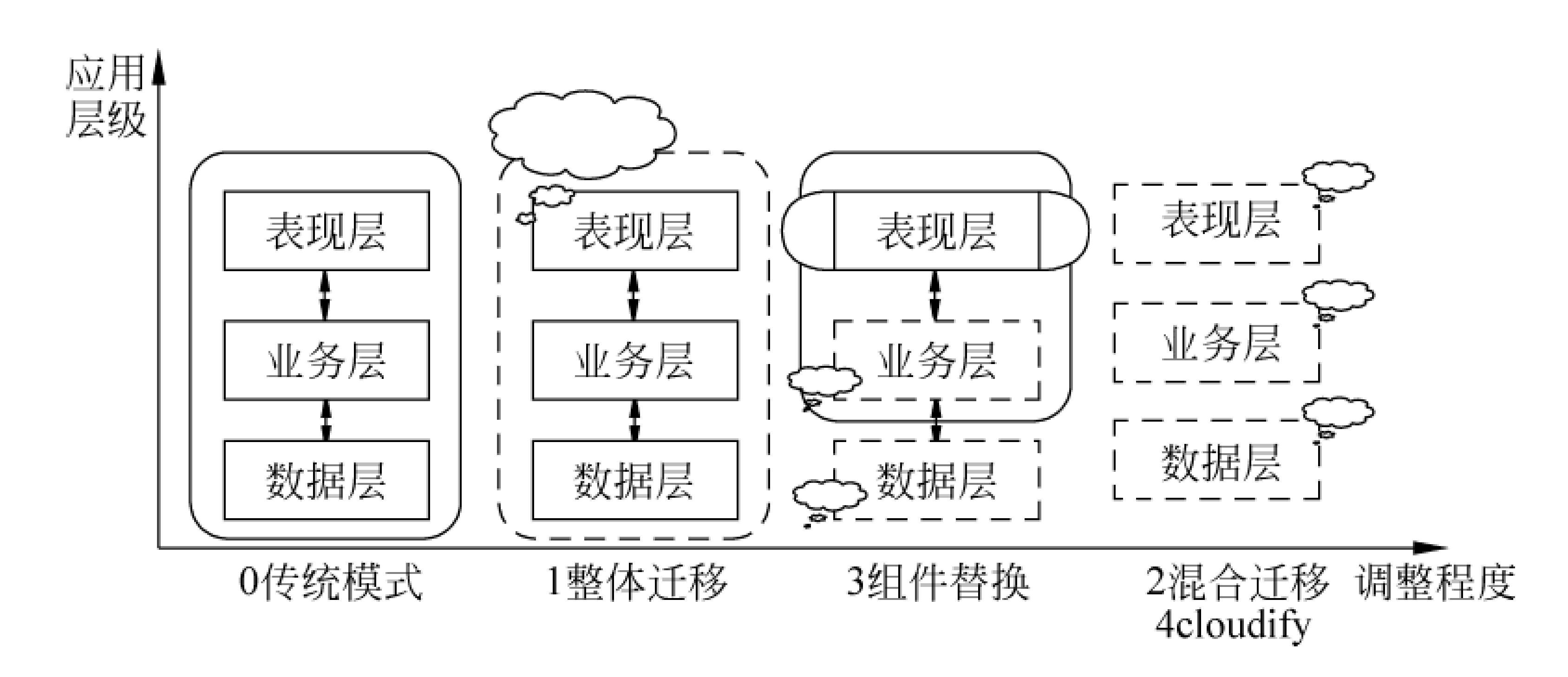


图 2-15　云迁移策略

的 IaaS 服务模型，核心技术手段是虚拟化技术。具体而言，应用程序的整个软件部分被封装在镜像中，以单一虚拟机的方式在云中运行。

在整体迁移策略下，应用程序在离线情况下被打包成镜像，以远程虚拟机的方式在云中运行。这种方式下，原有应用程序几乎不需要修改便可以在云环境中运行，这保证了对应用程序的最小入侵性，适用于小型应用程序。但这种迁移方式只利用了云计算的虚拟化特性，并没有发挥云应用的弹性、可扩展性等特性。

2. 混合迁移

混合迁移策略是指一部分应用或者组件迁移，另外一部分保留在原来的部署环境。现有的应用大都是多层式的软件架构，如图 2-9 所示的常见三层架构（表示层、业务层、数据层）的应用。多层级应用的各层间服务对象不同、处理的业务不同，因此对云资源的需求不同。混合迁移更多地关注软件内部架构，考虑将应用程序的某一个架构层而不是整个软件栈迁移到云平台上，而其余层的部件继续留在原来的物理环境中。更进一步而言，可以将一层或多层的多个架构组件迁移到云平台中，并且针对每一层的特点制定不同的部署策略。例如，可以将负载灵活变化的业务层迁移到云平台，为其提供弹性的资源供给，而把安全级别高的数据留在企业内部。

混合迁移策略更细致地考虑了应用程序与云环境的结合点，但它需要决策者决定哪些应用或应用层级应该被迁移至云平台。此外，如果迁移到的是公有云平台，则通信延迟也是需要考虑的问题。

3. 组件替换

组件替换策略是指利用云服务组件来替换软件的一个或多个组件。整体迁移立足于 IaaS 服务模型，采取基于虚拟化技术的迁移方法。混合迁移关注软件内部结构，但它对云服务的使用也仅限于 IaaS 层。然而随着云服务提供商开发出了越来越多的 PaaS 级或 SaaS 级服务产品，应用向云中的迁移方式便有了更多的选择，

可以使用云服务商提供的云服务组件来直接替换自身软件的一个或多个部件。这些云组件基于云环境设计开发并直接部署运行在 IaaS 云上，它们可以自动化弹性扩展、负载均衡、故障检测和处理，以及提供了多种级别的安全保障等，用户只需按需使用。例如，针对遗留系统，最可行的替换策略就是用云数据库替换其数据库。例如，Google App Engine Datastore 可替代本地 MySQL 数据库；Microsoft 的 SQL Azure 可以替换 SQL Server 数据库，Amazon 的 RDS 可替换 Mysql、Oracle 等多种关系型数据库。未来基于服务的应用在云中可以很方便地进行服务的替换和组装。

但组件替换很有可能造成新的云服务组件与原有组件之间的不兼容性，因此需要一系列的重新配置、重写或其他适配活动，对应用的入侵性较大。

4. 应用云服务化（cloud-native）

应用的云服务化是指将应用的所有组件由对应的云服务组件替换，这些云服务组件将组合成一个功能完整的云应用程序，最终运行在云环境中。相比较于混合迁移只迁移选定的组件，它将应用的迁移范围扩大到了应用程序的全部组件，是云服务程度最高的迁移方式。可以将应用横向划分为功能独立的子应用，从而选择 SaaS 级的软件服务进行替换，也可以将应用纵向分层，将每一层用相应的云服务替换。这种方式要求云服务提供商能够提供多种多样的云服务（IaaS 级的镜像—虚拟机服务、PaaS 级的基础支持服务、SaaS 级的软件功能服务）供用户选择。

应用云服务化后能完全地适应云环境，例如可以支持弹性扩展与负载均衡，但这种方式应用的重构程度和调整程度相比较于其他方式最高，而云服务组件与原有组件之间的不兼容性也需要考虑。

2.5.3 云部署策略

云部署是指从经济和易用性两个方面来合理规划应用对云资源的使用，以确保云环境能更好地为应用的正常运行提供服务。下面介绍几种实用性较高的云部署模型。

1. 单实例模型

单实例模型与整体迁移策略配合使用，是最简单的一种部署方式。这种模式不关心应用程序的内部结构，只是简单地将应用程序、数据、依赖环境所组成的迁移对象整体打包，制作成一个镜像，在云环境中启动一个虚拟机中来运行，如图 2-16 所示。由于数据库服务器与应用服务器绑定在同一个虚拟机中，为了保证数据的一致性，实例无法进行横向扩展，同时大数据量的情况下容易造成虚拟机瘫痪，因此这种模型适合数据量不大的小规模应用。当然，此模型支持纵向扩展，通过纵向改变实例的类型，可弹性地来提高管理实例的处理能力。

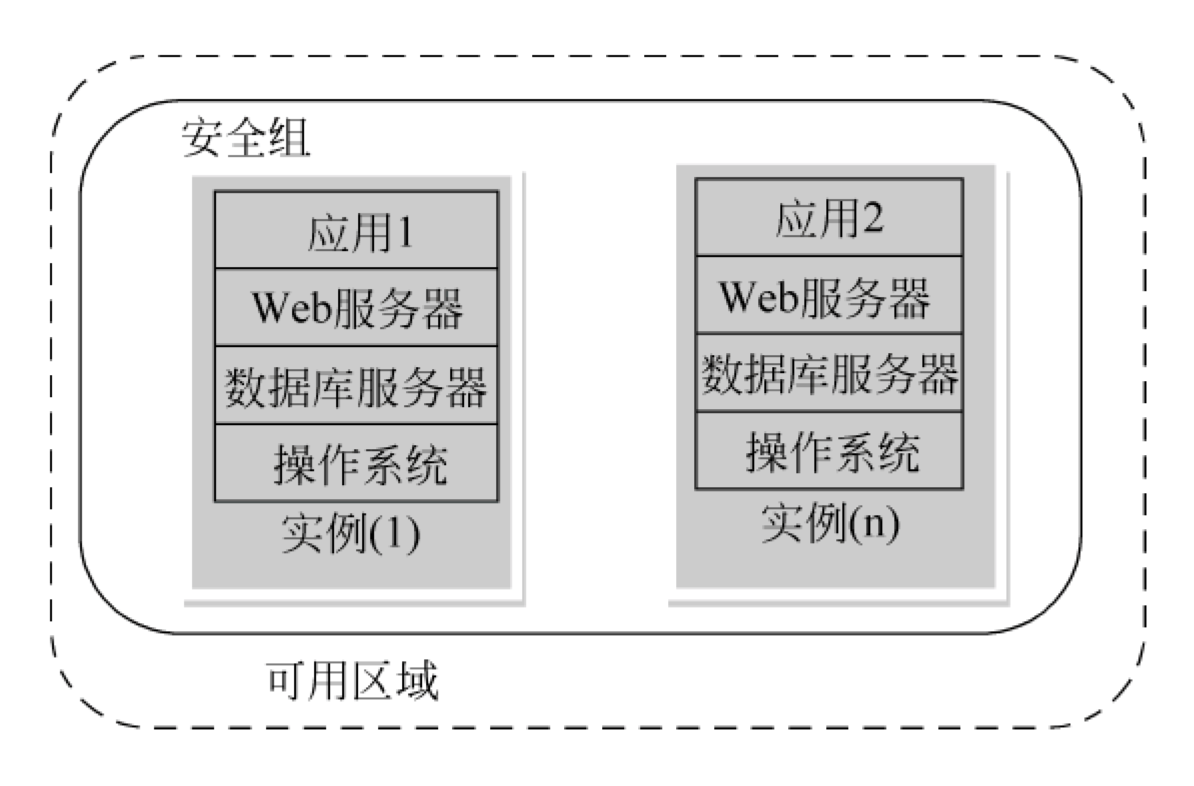


图 2-16　单实例模型

2. 应用与数据分离模型

多层级的应用程序，其不同层对硬件资源的使用需求不同，对安全控制的要求不同，弹性扩展的策略也不同。因此将需求不同的层装载进不同的镜像，启动不同的虚拟机来运行，能更加灵活地选用云资源。最常用的分层方法为应用与数据分离模型，如图 2-17 所示。首先需要将数据库服务器从单实例模型中分离出来，单独创建数据库服务器实例。应用程序和 Web 应用服务器以及操作系统放置在一个实例中，作为弹性可扩展的最小单位，并可为该应用实例组配置负载均衡服务，它负责将对同一个应用的请求均匀地分发到各个实例上去。数据库实例也可以有其自己的云资源使用模式和弹性扩展策略。

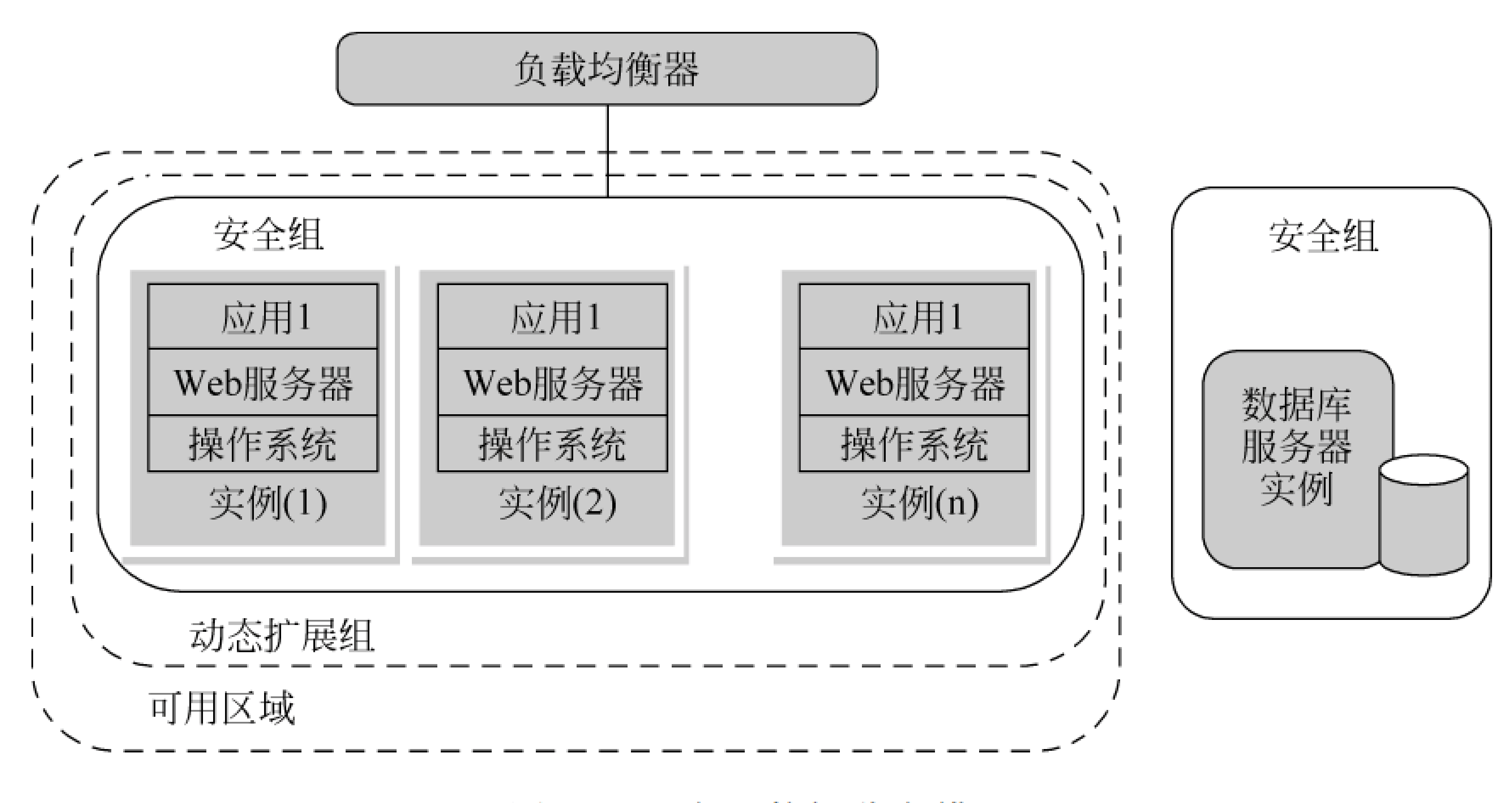


图 2-17　应用数据分离模型

应用服务属于计算密集型，数据库服务属于 I/O 密集型，因此可以为它们选择不同的实例类型。此外可以为这两种实例创建不同的弹性扩展组，实施不同的弹性扩展规则。这种弹性扩展包括横向扩展和纵向扩展两个方面。弹

性扩展应用实例可更好的应对来自客户端的请求量的变化，与此同时，数据库实例也可随着存取访问量的变化而适当地弹性扩展，从而保证整个 Web 应用的正常运行。

安全组定义了一系列针对组内实例的访问规则，可有效控制来自外围的访问，为应用服务提供安全保障。可以根据需求为两种实例创建不同的安全组，从而更加严格地控制对实例的访问。

3. 业务的拆分与组合

调整应用程序的业务组件，从时间和空间上协同使用虚拟资源，可大大提高资源利用率，节省运营成本。一般从以下三个方面实施：

1）任务的并行粒度

对并行程度高的应用，可以调整并行模式以选择不同的资源：选择一个包含多核 CPU 的计算能力强的虚拟机；选择多个单核微型虚拟化进行集群式并行处理。计算这两种模式下的经济投入，从而择优选择。

2）业务整合

整体硬件投资不变的情况下，可以通过错峰运行调度业务压力，即将访问峰值不在同一个时间段的业务整合在同一个虚拟机中。

3）灵活规划扩展策略

例如，业务时间使用大容量虚拟机，闲时使用小容量虚拟机；或者测试业务的瓶颈，在高峰期横向扩展瓶颈业务等。

2.5.4 云迁移生命周期

图 2-18 描述了云迁移工程的整个生命周期，下面详细介绍每一阶段。

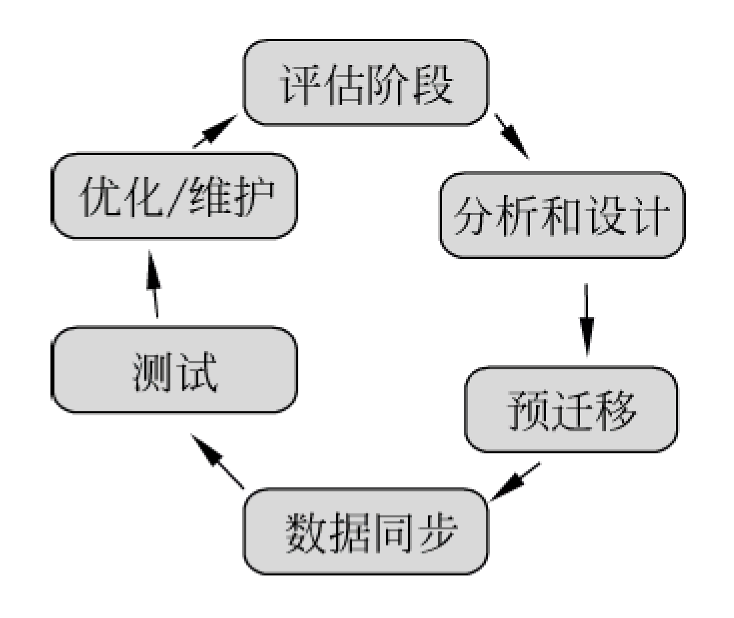


图 2-18 云迁移模型

1. 评估阶段

云迁移的评估阶段需要收集的信息包括项目管理信息、迁移的动机、迁移的目标云平台、需要的工具等等。具体来说，包括以下核心内容。

1）迁移考虑的动机

（1）节约运营成本。

云计算采用按需付费的服务方式，这意味着购买资源的支出与当前的需求动态匹配，而不需要在前期就计划购买足够支撑业务峰值的设备。因此，节约运营成本就成为了企业选择云迁移工程的首要动机。

（2）提高应用程序的可扩展性。

可扩展性是指系统通过增加资源来适应更大的负载的能力，常见的增加资源的方式有两种：一种是横向扩展（scale out），即添加更多的资源节点；另一种是纵向扩展（scale up），节点数不变，使当前节点的硬件资源更强，计算存储等能力倍增。

（3）提高资源的有效利用率。

传统模式下通常一台服务器上运行一种服务，大多情况下资源处于闲置状态。云计算采用虚拟化方式，支持一台物理机、多台虚拟机运行多种服务。此外，根据业务访问的特点，将不同访问高峰的业务整合到一起，从时间上划分对资源的使用，更能大大提高资源的利用率。

（4）弹性。

弹性是指调整资源需求以适应动态负载变化的能力，这通常与横向扩展关联。这样当负载增加时可以增加资源来扩展，而当需求减弱时退缩并删除不需要的资源。弹性是云计算环境的一大特性，对计算资源使用适时动态变化以应对变化的动态负载。

（5）可维护性。

在云计算环境下，应用程序以镜像—虚拟机的形式运行，在运行过程中也可实时创建快照保存运行时状态。因此，当应用访问出现故障或者硬件出现故障时，可以以镜像为模板重新启动一个实例，快速地恢复运行。

不同的迁移动机意味着在迁移时要重点考虑不同的问题。例如，如果希望云中应用能有较好的弹性、可扩展性，以灵活应对负载变化，在迁移之前就要考虑应用程序的软件架构是否可以扩展，同时调整应用程序在云环境中的架构，使它的可扩展性达到最佳。如果要提高资源的利用率，就要分析应用程序对资源的需求模型，设计出合理的资源使用模型。如果要节约运营成本，就要再结合云服务提供商对云资源的定价进行分析，构建经济模型。

2）当前应用环境的清单

创建当前应用环境的详细清单有助于理解迁移工程所映射的范围，如图 2-19所示，清单包括程序数量、编程语言、部署平台、第三方组件或框架、脚本、外部接口、最低硬件配置、软件配置等信息。比如操作系统版本、数据库版本、使用中的应

用特性(QoS 要求)或功能,以及其他类似的信息。

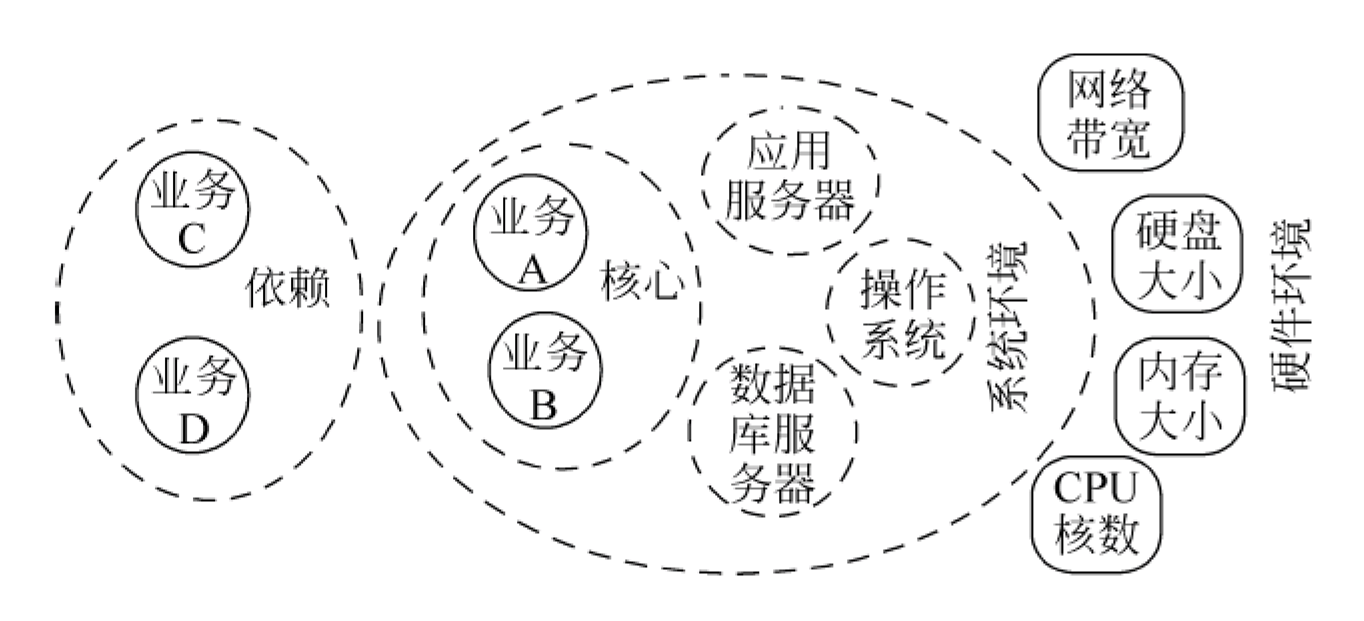


图 2-19 当前应用环境清单

3）云服务及迁移工具

选择合适的云服务提供商是迁移工作的重要环节。用户首先需要明确选择公有云服务、私有云服务、还是公有和私有共同使用的混合云方式。确定云服务方式后需要对常见的云服务提供商提供的服务进行评估,包括服务模型、提供的资源(虚拟机、存储空间、部署环境)、资源的成本、安全机制等。

云环境的易用性也是需要考量的方面,例如是否提供强大的迁移或部署工具、是否有易用的编程接口和友好的访问界面等。

2. 分析和设计

在评估阶段,对软件环境进行系统性分析,确定待迁移至云的应用程序。分析和设计阶段需要对待迁移的应用进行详细分析,最终确定应用在云中运行的诸多细节。此阶段涉及两大核心任务：一是根据评估阶段收集的软件环境清单,分析现有应用的整个软件环境,提取出现有的软件总体架构；分析自身业务需求,选择云迁移策略,设计出应用迁移至云环境后的目标软件架构；二是分析现有环境的硬件资源,检测应用的资源使用情况,结合云供应商提供的基础设施资源,确定预迁移至云的应用的云资源需求。

1）基于云环境的目标软件架构

(1）提取应用的软件架构。

架构就是指系统组成元素及各元素之间的关系,云迁移工程的实施对象是正在运行的应用,相对于软件工程来说,它是一个逆向工程,即从现有的应用中提取出软件体系架构。

早期的应用系统主要是单机软件系统,多数行业将软件技术当做辅助手段来解决自己专业领域的问题,其中大多是较深入的数学问题或图形图像处理算法的实现。从软件架构的角度看,单机系统是比较简单的,如果要将此类型应用迁移至云环境,研究的重点是提高应用程序的并行处理能力。无论是 CPU 级的并行还是虚拟机级的并行,云环境都能按需提供资源。实践经验表明,云迁移的主要目标群

体在信息系统领域，即以数据处理(数据存储、传输、安全、查询、展示等)为核心的软件系统。

从云迁移的角度来看，常规信息系统的软件架构如图 2-20 所示，由下至上分为三层：第一层为数据层，第二层为业务层，第三层为表示层。其中数据层可分为数据库层，包括常见的关系型数据库、非关系型数据库或其他类型的数据库。业务层可细化为业务逻辑、应用服务器、中间件/基础服务层。

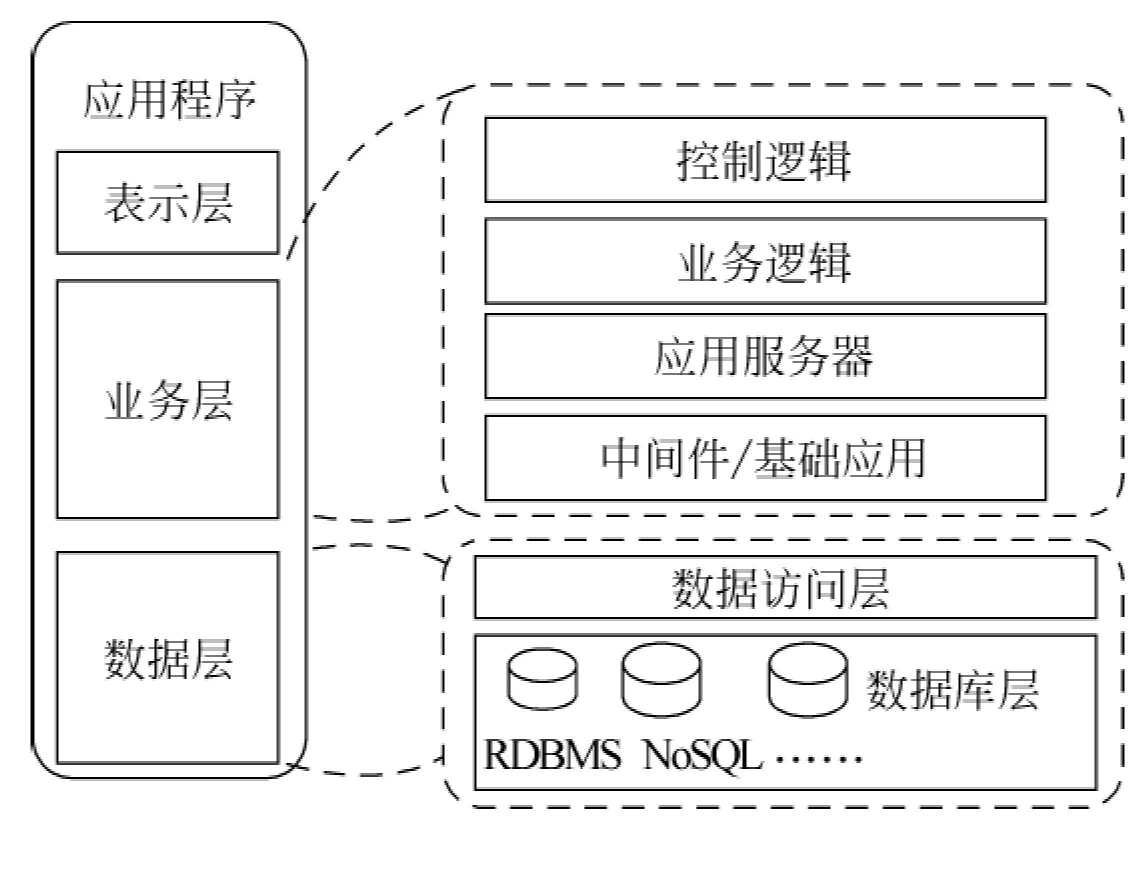


图 2-20　源软件架构

(2) 目标软件架构。

云迁移将应用程序的每一独立层看成一个黑盒而不关心其内部的实现原理或逻辑结构。选择什么样的迁移策略和部署策略，决定最终的应用的目标软件架构。如果选择整体迁移策略、单实例模型，则目标软件被打包成一个镜像在一个虚拟机中运行，其软件架构并没有发生变化。因此，需要考虑在其他迁移策略下，目标软件架构所发生的变化。

图 2-21 阐释了多种迁移策略实施之后的目标软件架构的多种可能性。如果将应用完全云服务化，则其整个软件架构如图 2-21 中浅色覆盖的区域。首先，采用数据库分离的部署模型，应用程序和数据库分别部署于不同的实例中。而为了减少云环境中频繁业务访问带来的带宽压力及网络访问，需要对云应用做两级的缓存处理：一级是应用层的静态缓存，另一级是数据层的数据库缓存。其次，无论是应用层、数据层还是两级缓存，为了应对大容量访问，都被设计成可扩展的多集群式处理模式。最后，云应用与最后用户之间要再加一层防火墙保护一级负载均衡服务，从安全和性能两方面来保护云应用。这样的云软件架构设计，应用可以充分利用云环境相对于传统环境的弹性、可扩展性、易维护性等优势，更好地在云环境中运行。

出于安全、经济、性能等原因，用户可能并不想把应用程序全部搬到云平台上

来，而只想采用混合迁移的策略选择部分业务迁移。例如，用户希望核心数据留在自己的防火墙内部，而只把应用程序迁移至云平台运行。在图 2-21 中，用深色区域的数据层来代替红色虚线部分的数据层和数据库缓存层。这种架构下，最终用户访问的请求将由运行在云平台上的应用层来处理，而业务处理所需要的数据访问则要发回运行在企业内部物理服务器上的数据层处理。如果选择公有云服务，就要充分考虑到数据访问请求的网络延迟对应用性能的影响，延迟敏感型应用则不应该被设计成这种架构。如果是在私有云的模式下，则要使数据库服务器与运行应用程序的云节点处于同一局域网，并保证二者之间通信链接的带宽能力。

如果使用组件替换策略，则可以考虑将应用的一个独立部分托管给云服务提供者。如图 2-21 中用云数据库服务（如 Amazon RDS、Google APP Engine Datastore）来替换数据层，那么整个数据层将全部托管给了云服务提供商，用户只需按需付费即可。

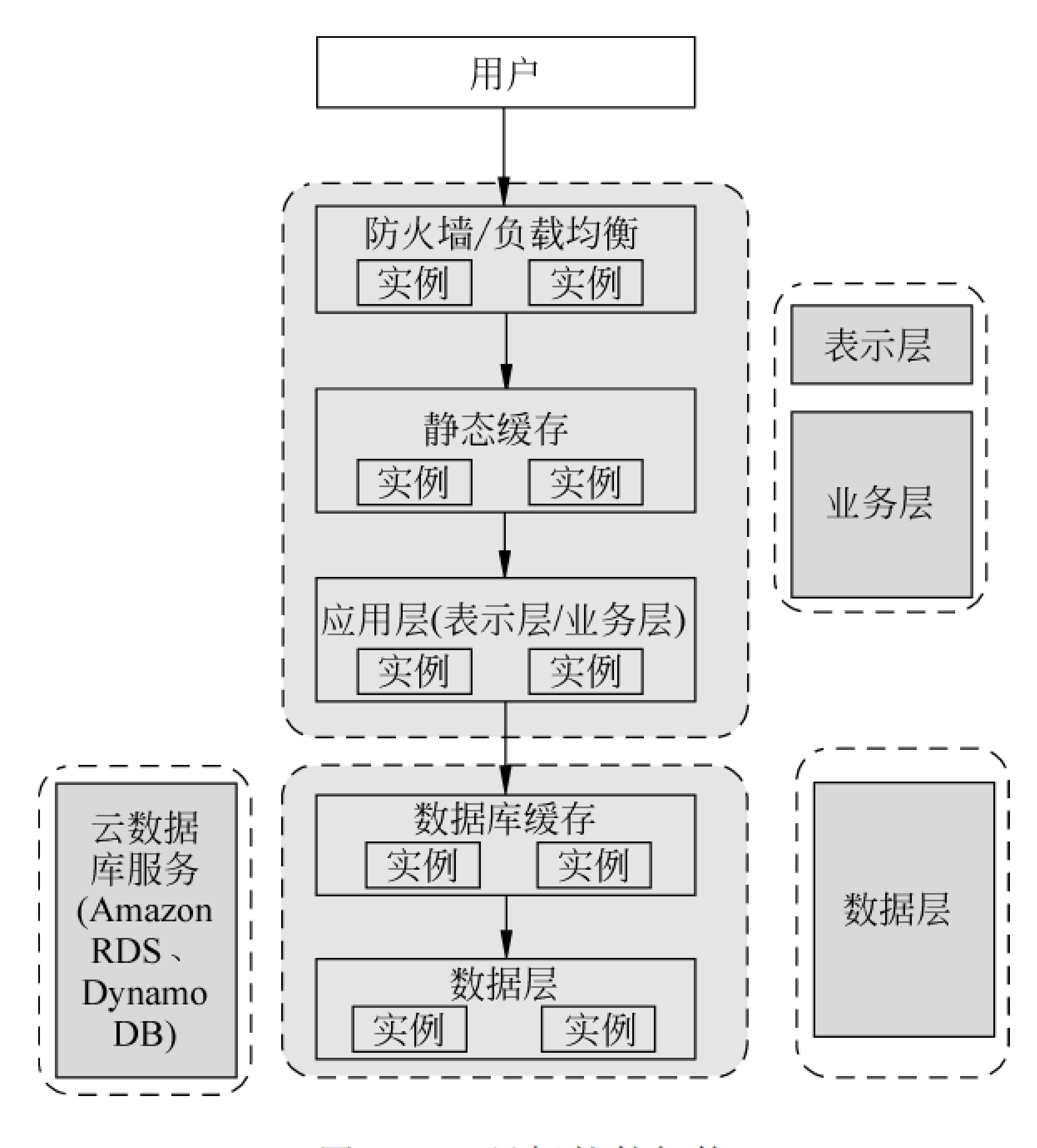


图 2-21 目标软件架构

上面对三种迁移策略下可能的目标软件架构做了分析。在实际的工程项目中，要根据具体的应用做具体分析，灵活应用。

2）应用的云资源需求

IaaS 级云服务按需提供基础设施服务，因此云迁移设计阶段需要仔细分析应用对基础设施资源的需求，确定最终的云资源使用方案。

（1）了解现有应用的资源需求。

这需从两个方面考虑对资源的需求：一是现有应用的硬件资源配置，主要包括CPU配置、内存配置和存储配置；二是对应用运行期间的性能进行测试，得出与性能相关的参数，如平均负载、峰值负载、安全系数和业务增长预测。这些数据的得出通常要借助于第三方监测工具或者经验丰富的专业人士。

（2）确定部署与扩展方案。

前述两步将原始资源需求转换为目标资源，第三步则应确定最终的部署和扩展方案，即采用相关的云部署策略，从经济和可扩展性两个方面来合理规划应用对云资源的使用。

• 部署方案。

根据之前确定的各项云资源的配置参数，并选择一种云部署策略来确定最终的应用部署方案。选择哪一种云部署策略的关键在于预先确定的部署目标。例如是要最大化的保证应用的可扩展性以应对灵活多变的业务需求，还是要尽可能提高基础设施资源的利用率（私有云），或降低IT业务的运营成本（公有云）。目标不同则部署策略的选择不同。

• 扩展方案。

扩展方案包括两个方面：负载监控和扩展策略。负载监控是指在承载自身各组件的虚拟机内部，对各项负载（包括CPU、内存、磁盘I/O）进行监控。扩展策略由用户根据自身业务访问特点制定，例如当某个参数超过某个阈值时，动态增加一台虚拟机；当某个参数低于某个阈值时，动态减少一台虚拟机。增加和减少虚拟机的过程如图2-22所示。当监控的参数到达某个阈值时，根据扩展策略，用户可以使用装有应用程序的镜像模板，启动同等规格的虚拟机，再与其他虚拟机同步数据，并向负载均衡器报备。整个过程可以是用户手动完成，也可以采用云服务提供商的动态扩展服务完成。

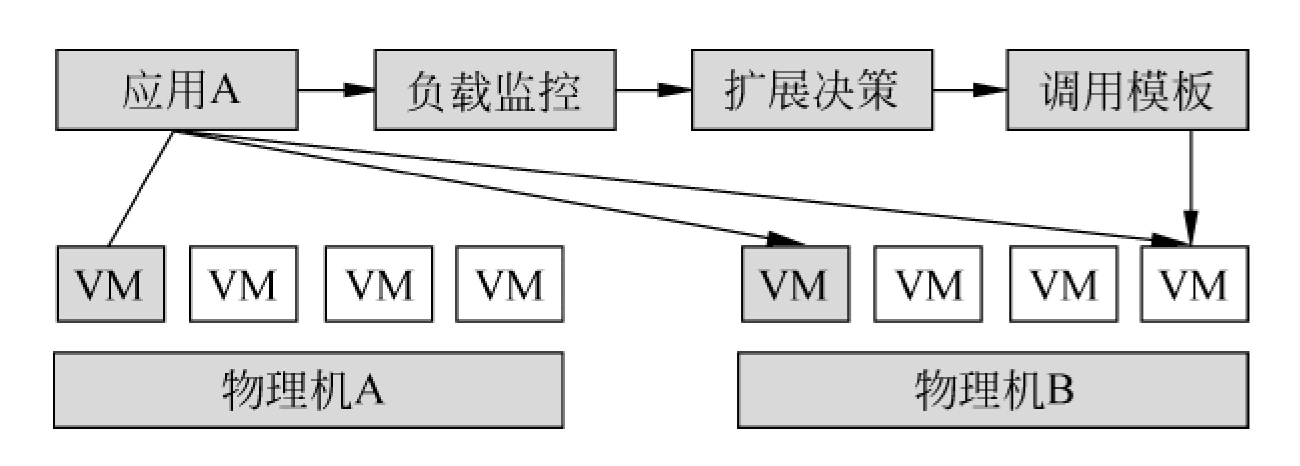


图2-22　扩展策略

（3）确定虚拟机的配置。

根据确定的应用资源需求和部署方案来确定最终的云资源配置，这里主要指虚拟机以及存储的配置。对虚拟机的配置选择应遵循以下原则：

① 单台虚拟机能够应对日常的业务流量；

② 所有的虚拟机尽可能统一配置(CPU 和内存配置);

③ 在预算相同的情况下,当对性能要求较高时选择多核、大内存的虚拟机;当对程序可用性要求较高时应该选择多个虚拟机的解决方案。

3. 预迁移

此阶段根据选择的迁移策略,以及最终的目标软件架构来执行应用到云环境的实际迁移活动。这些活动包括迁移所需的任何具体活动:将应用程序组件和数据移动到所选择的云服务平台,并执行所有的变化和必要的测试来确保应用在云中按预期的状态运行。

1) 申请云资源、启动云环境

启动云环境的第一步是获取云资源的使用权限,这包括注册账户、获取证书、生成密钥以及资源的租用费等。获取权限之后就可以配置云资源环境,包括虚拟机类型、镜像类型、网络配置、安全配置、存储配置等等。之后便可启动实例,调试网络、存储、软件环境,使实例正常运行并可被外界正常访问,具体流程如图 2-23 所示。



图 2-23 启动云环境

2) 应用程序迁移

应用程序迁移的流程如图 2-24 所示。首先登录到实例搭建应用运行所需的全部软件环境(OS、语言环境、数据库、Web 服务器等系统软件),运行时环境搭建完成后便可将应用程序从原环境中打包部署到新的软件环境。如果使用部分迁移或者组件替换策略,则需要根据设计好的目标软件架构来调整应用程序,必要时要做些代码上的修改。此外,部署应用程序的实例最终要被作为动态扩展的模板,因此在部署成功后要将当前实例保存成镜像,并妥善存储。



图 2-24 应用程序迁移

3) 样本数据迁移

云环境启动,并且应用程序及其软件运行时环境部署完成后,接着要迁移部分的业务数据来试运行业务,这部分数据被称为样本数据。

数据的迁移多数情况下是指数据库的迁移,而信息系统的数据库多数都是关系型数据库。关系型数据库迁移至云环境主要有两种方式:

(1) 采用云数据库服务替代自身的数据库服务,将数据管理托管给云服务

平台；

（2）自己管理云环境中的数据库服务。如果是传统集中式的数据库，数据库向云迁移主要是因为其扩展性，传统集中式的数据库需要转换成分布式可支持动态扩展的数据库。

如图 2-25 所示，为了能与用户的源数据库环境保持一致，数据库的迁移要考虑的内容有：

（1）数据库模式 schema：表、索引、视图等；

（2）数据库存储过程：存储过程、触发器、视图；

（3）数据库管理脚本；

（4）数据库的数据。

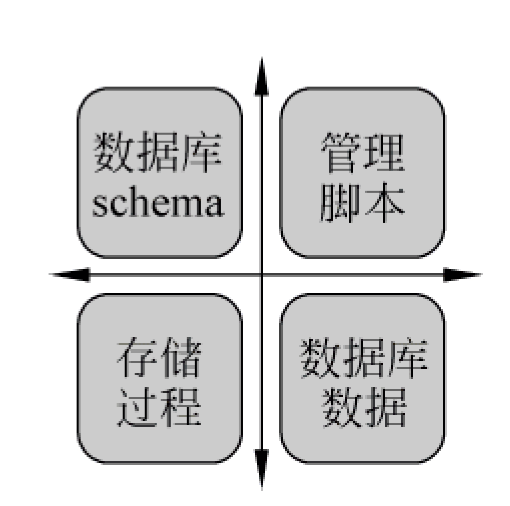


图 2-25　数据库迁移

最终在云环境中建立起与源数据库环境一致的目标数据库，并且在其中加载样本数据。

4）启动服务

等数据就位后，就可以启动业务相关的各项服务来调试云应用程序。这里要注意对应用程序各个独立模块间的部分配置进行修改，如修改应用程序中数据库链接的 IP。

5）预测试

服务启动后，可在样本数据的支撑下调试云应用的试运行环境，保证服务正常运行，并且从外界可正常访问。

4．数据同步

数据同步即将源数据库中的数据迁移到已在云中正常运行的云数据库中。迁移数据的一般流程是使用脚本调用数据库工具来从源数据库提取数据，再导入到目标数据库中。具体迁移数据的方法包括离线迁移、在线迁移和追踪数据变更。

1）离线迁移

源数据库和目标数据库间并不建立链接，数据先从源数据库中被提取出来，以文件的形式保存，再在目标数据库中使用本地工具或脚本在空闲时间加载该数据文件。

2）在线迁移

使用 JDBC 驱动、ODBC 驱动、第三方数据库管理工具将源数据库和目标数据库链接起来，在它们之间直接迁移数据，在此期间要保持网络的可连接性。这种模式的数据迁移会为源数据库所在系统带来额外的负载开销（CPU、内存、I/O 操作），因此不建议在大型数据库或任务量重的数据库上使用，它适用于小型数据库且工作负载不太大的情况。

3）追踪数据变更

记录发生在源数据库的变化，并阶段性地复制这种变化到目标数据库。这是大型的、不可间断数据库迁移的有效手段。

实现方法有两种：

(1) 日志挖掘，即在源数据库中读取已归档的事务日志，提取其中执行了的事务操作，随后将其保存成中间文档传输到目标数据库或者直接传输并在目标数据库中执行。例如 MySQL 数据库使用 binlog 技术，批量读取源数据库的 binlog，一条一条同步到目标数据库，执行相应的数据库操作。

(2) 触发器，即在源数据库中设置触发器，捕捉该数据库上的变化并将其写入临时表，再将这些变化从临时表复制到目标数据库。

5. 测试

此阶段需要对前几个阶段的迁移成果做定量、定性验证，这主要从功能和性能两个方面进行。

1）功能测试

功能测试以黑盒方式测试应用程序迁移到云环境后其各项业务功能有无发生损耗或出现障碍。

(1) 数据库迁移的测试。

确保数据服务在云中正常运行是测试任务的重中之重，测试的内容包括：

① 数据完整性确认——数据大小是否一致，编码方式是否设置正确；

② 数据库存储过程是否能正常运行；

③ 数据库维护脚本(备份或恢复)还能否发挥作用等；

④ 数据库模式是否正常运行。

(2) 应用测试。

应用测试包括两方面：一是测试应用程序运行在各个虚拟机中的独立模块间能否正常通信；二是从最终用户的角度测试各项业务功能，验证其服务是否正常。

2）性能测试

性能测试考量应用与云环境相互合作的默契度。一是要测试云环境，确保云资源是真的按需提供，以保障迁移者的利益；二是要测试云应用服务性能，确保最终用户的用户体验。

(1) 云环境测试：可以借助第三方工具探测虚拟资源配置是否按预设需求提供；

(2) 云应用测试：测试应用的性能是否在计划范围之内。

6. 优化和维护

1）弹性扩展和负载均衡

大多数云服务平台都提供弹性扩展服务和负载均衡服务，开发者可以按照设

计阶段设计好的扩展方案实施。但云服务平台提供的弹性扩展一般是虚拟机级别的扩展,不涉及具体的应用,因此虚拟机扩展后,需要用户自己去同步服务。同时,用户也可分别为每一层的服务做单独的负载均衡。

2）高可用性

为保证云应用可以无障碍不间断服务,可以在关键性业务上启动备份实例,同时交叉部署负载均衡器管理的各个服务实例,使得底层设施出现故障时不会影响到上层服务,从而保证云应用的高可用性。

3）备份、恢复

从应用程序以及云资源两个方面去实施备份、恢复策略。应用程序方面可以沿用传统环境时所启用的那些管理方案,而云资源方面则要采用服务商提供的工具,定时备份虚拟机和虚拟磁盘,保证资源使用的完备性。

2.6　云计算与大数据在省市级社区矫正信息系统中的应用

2.6.1　概述

2014 年 4 月 22 日,中央政法委书记孟建柱在北京出席第一期政法领导干部学习贯彻习近平总书记重要讲话精神专题培训班开班式上指出:“要善于运用大数据,提高维护稳定工作现代化水平。谁率先拥有、善于利用大数据,谁就能掌握主动、赢得未来。”可以说,对于司法行政机关来说,大数据时代的到来,既是机遇也是挑战。

社区矫正是指将社区矫正对象置于社区内,由专门的国家机关负责并组织社会力量对其采取监督管理、教育、帮助措施,矫正其犯罪心理和行为恶习,促进其顺利回归社会的非监禁刑罚执行活动。随着信息技术的发展,社区矫正信息化进入了一个新的阶段,特别是云计算和大数据技术的发展和广泛应用,为社区矫正信息系统提供了基础支撑环境和平台。

《社区矫正实施办法》第 35 条规定:司法行政机关和公安机关、人民检察院、人民法院建立社区矫正人员的信息交换平台,实现社区矫正工作动态数据共享;实现公、检、法、司等各个部门之间的联网,加强部门之间的横向联系,信息互通,资源共享。社区矫正数据库包括专家库、案例库、标准库,其中专家库主要是在社区矫正领域最前沿的专家学者的理论研究和对实证的评论;案例库主要是司法行政实务部门海量的社区矫正案例和工作的经验、技巧;标准库主要是经被证明其经验具有较大可靠性的经典法律资源,例如,被矫正人员在解除社区矫正多年以后未再违法犯罪的矫正经验。经过分类整理,数据库使用者可以在最短的时间内准确

地选择检索范围，并且方便、快捷地获取所需要的研究资源或社区矫正工作所需要的案例指导，为社区矫正实践奠定坚实的基础。要建立社区矫正数据库，不但需要上下级之间的指导与配合，还需要联系社区矫正参与各方提供数据与案例，联合与发动社区矫正研究者提供理论成果，甚至应成立社区矫正社会调查组织，来调查被矫正者矫正前、矫正中、矫正后的情况，所以社区矫正数据库的建立是一个庞大的系统工程。

云计算与大数据正在改变司法行政信息化及社区矫正实践现状并成为应对新时期司法行政工作突破瓶颈的新方式、新视角、新路径。在云计算和大数据时代背景下创新社区矫正工作，在更高的起点上推动社区矫正工作实现新突破，具有重大的现实意义。

2.6.2 系统框架

社区矫正工作主要包括制订矫正工作计划、接收矫正对象、制定教育矫正方案，以及对矫正对象实施法制教育、公益劳动和培训学习等日常的监督管理。根据矫正对象的考核情况，按照规定提请对其进行行政奖惩、完成上级社区矫正领导机构布置的相关工作。主要业务包括矫正观察、档案管理、教育矫正、考核管理、信息管理等主要业务。

目前我国在大力推行社区矫正工作机制，部分省市在应用信息化手段开展社区矫正工作方面，取得了一定的成果。但是普遍存在单项应用较多、综合应用少、信息资源缺乏统筹开发、共享率低等问题。社区矫正工作涉及省、市、区等多级管理机构，具有矫正对象数量多、采集信息量大等特点，因此，需要一个强大的业务服务支撑平台。现在，云计算和大数据技术的最新发展为解决这一难题提供了一种全新的、完整的体系架构，它对于社区矫正信息系统的运行以及大数据处理可以起到非常重要的支撑作用。通过云平台可以从纵向角度，整合省、市、县区及社区各级司法行政机构的软硬件基础设施，建设统一的网络、数据中心等硬件支撑平台；从横向角度可以整合各级司法行政管理机构的信息资源，做到资源共享，提高信息利用水平。

基于云计算和大数据的省市社区矫正信息系统框架如图 2-26 所示。

1. 基础设施服务

IaaS 层构建整个平台的硬件基础，包括各种服务器、存储设备、网络设备（如路由器、负载均衡设备、防火墙）等，实现硬件资源的虚拟化聚合管理，为各类应用系统提供运行和数据存储服务。通过硬件资源的虚拟化聚合为计算资源池、存储资源池和网络资源池，以虚拟桌面的方式为用户提供云服务，或通过资源分发等接口为 PaaS 层等上层应用提供硬件资源服务。

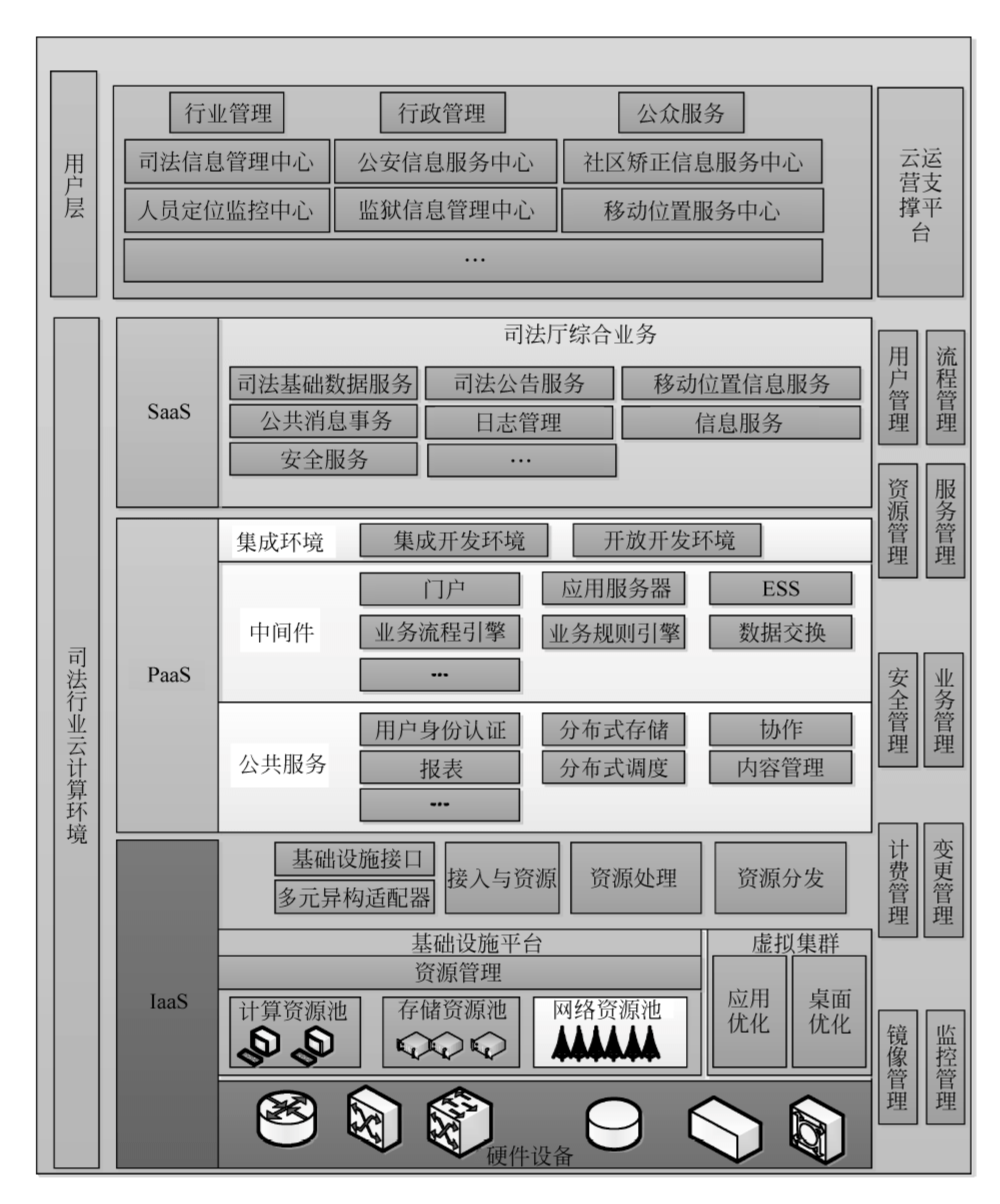


图 2-26　社区矫正云计算平台框架

2. 应用平台服务

PaaS 层提供整个平台的应用系统开发和运行环境，包括集成环境、中间件和公共服务等，为各类司法行政及社区矫正信息系统提供运行环境和数据库等支撑环境。其集成环境主要提供应用程序开发和运行环境，中间件主要提供企业服务总线（Enterprise Service Bus，ESB）等应用程序开发和运行的各类通用中间件，公

共服务主要提供用户身份认证等应用程序开发和应用的基础模块，例如面向大数据处理的分布式存储等模块。

3. 应用软件服务

SaaS 层为平台提供用户最终所需要的业务服务功能，包括区域监管（包括实时定位、历史轨迹、电子围墙）、信息交互、网上办公、信息发布、教育帮助、考核鉴定、审前评估、假释评估、安置帮教、风险评估、统计分析、工作机制等功能。通过这些服务功能，司法行政管理人员可随时随地了解矫正对象的位置（包括实时定位和随机位置查询），在预定时间内了解社区矫正对象是否离开安全活动范围，并提供考核依据。同时，该平台还可实现对社区矫正人员的集中学习、参加公益劳动、汇报思想、定期统计矫正工作数据等功能。

4. 用户层

为各类用户（省司法厅、市司法局、县区司法局、乡镇街道司法所）提供使用平台服务的各种方式，除了传统的 Web 浏览器接入门户外，还可提供智能手机等移动终端接入门户。为社区矫正主管部门及其他政府部门提供行业监管、政府决策、应急处置等应用服务。

5. 运行支撑平台

该平台提供整个平台的管理功能，包括用户管理、资源管理、流程管理、安全管理和计费管理等，实现社区矫正云平台"整体合规、资源可控、数据可信、持续发展"的管理服务与运行维护的目的。



第3章

基于云计算与物联网的社区矫正信息系统体系结构

社区矫正的主要目的是监控矫正对象、管理矫正对象信息及矫正的各项事务。我国已于2009年在全国范围内全面实行社区矫正制度。社区矫正管理信息系统不是一个孤立的系统,而是多个子系统的集合,各个子系统协同工作,共同完成社区服刑人员的教育矫正工作。因此,社区矫正管理信息系统需要强大的计算机网络、物联网技术和云计算及大数据技术的支持。

3.1　总体架构

社区矫正信息系统是利用物联网技术、LBS(基于位置的服务,Location Based Service)技术、计算机网络及云计算等技术所构建的信息平台,旨在将创新管理、结构优化、资源整合融入社区矫正的信息化建设中,为社区矫正各级管理部门及管理者提供统一的平台,提高司法行政信息化水平。采用物联技术可以全面管理、识别与定位矫正对象,而云计算是社区矫正信息系统的基础,提供大规模、分布式计算资源。基于云计算与物联网的社区矫正系统总体架构如图3-1所示。

物联网云的系统架构主要包含物联网云的硬件虚拟化框架、服务管理、物联网应用中间件以及感知层设备等。各部分共同构成物联网应用的平台,为物联网应用的运行和终端用户服务。各层基本功能如下:

1. 感知层

感知层是主要用于识别物体、采集信息。感知层设备包括二维标签码和识读器、RFID标签和读写器、传感器终端以及实现终端互联互通的传感网络。感知设

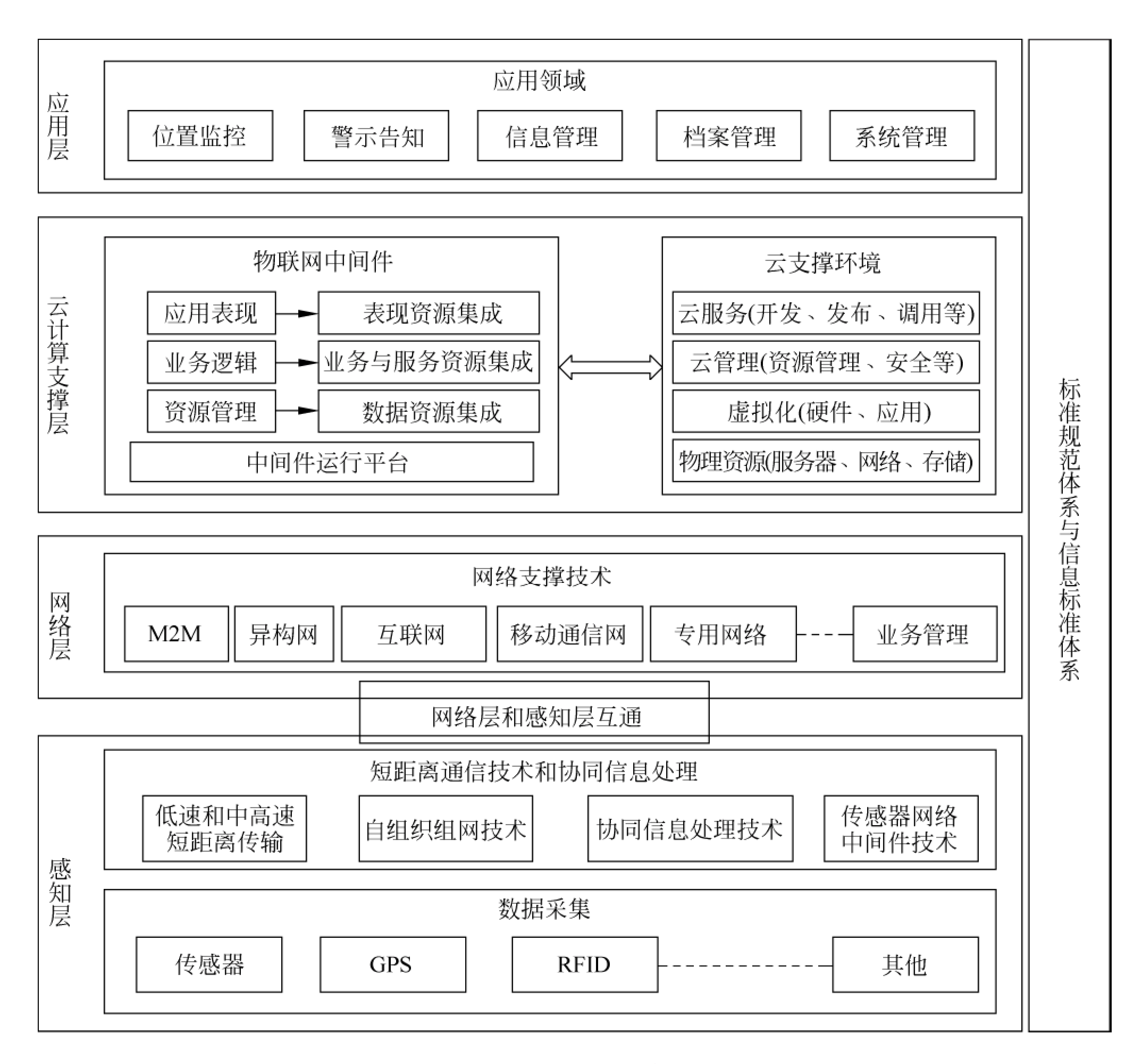


图 3-1　物联网云平台体系架构

备通过网络接入云计算平台，并由物联网应用的中间件对其进行管理。通过感知层设备，物联网可以给物体赋予“智能”，实现对物体的感知、人与物体的沟通和对话，也可以实现物体与物体间的沟通和对话。感知层涉及的关键技术有射频技术(RFID)、传感网络技术、智能嵌入技术等。

2. 网络层

网络层主要用于前端感知数据的传输，最大利用现有网络条件，融合司法专网、电子政务专网等。

3. 支撑层

支撑层基于云计算、云存储技术设计，实现分散资源的集中管理以及集中资源的分散服务，支撑高效海量数据的存储与处理；支撑软件系统部署在运行平台之

上,实现各类感知资源的规范接入、整合、交换与存储,实现各类感知设备的基础信息管理,实现感知信息资源目录发布与同步,为感知设备的信息处理、共享提供全面支撑服务。

4. 中间件

中间件是位于平台(硬件和操作系统)和应用之间的通用服务,针对不同的操作系统和硬件平台,它们可以有符合接口和协议规范的多种实现。中间件是物联网应用中的关键软件部件,是衔接相关硬件设备和业务应用的桥梁,其主要功能是屏蔽异构性、实现互操作和信息的预处理等。在物联网云平台中,物联网中间件与云计算相结合,利用虚拟化技术全面实现资源整合,这样,不仅能解决物联网中海量信息的过滤、整合、存储问题,还能解决物联网中不同应用系统的互操作问题。在物联网云平台中,物联网应用的中间件主要实现终端设备接入、RFID/传感器事件管理、数据存储以及物联网应用等功能。

基于拥有的丰富的数据资源和强大的计算能力(依托云计算平台),构建一个功能丰富的物联网中间件平台;基于SOA架构为应用层的实际业务应用软件提供统一的服务接口,对数据进行了统一高效的调用,保证服务的高可靠性,也为整个平台的后续应用开发提供可扩展性。支撑层对各数据标准的融合起到关键作用,是整个物联网运行的关键所在。

5. 应用层

应用层主要用来承载用户实际使用的各种业务软件,如位置监控、警示告知、信息管理、档案管理、系统管理等。物联网云平台不可能是一个封闭自运行的应用系统,需要具备第三方行业应用的集成能力,即要能提供给第三方合作开发者灵活拓展的云端应用开发API接口,从而能够满足不同应用的差异化功能要求。

6. 安全与保障体系

安全体系设计主要是从系统的安全威胁和风险分析,明确安全等级保护实施办法以及与保护等级相适应的安全策略,以实效和应用为主导,管理与技术并重,从物理、网络、系统层、信息交换层、应用层、组织管理等方面,保障系统的安全,满足系统的建设和发展要求。

信息标准建设是系统实现互联互通、信息共享、业务协同的前提和基础。早期的信息系统建设,普遍存在缺乏标准或标准不统一的现象,最终导致管理混乱、互联互通不畅、信息共享程度低、信息资源开发利用滞后等后果,严重影响行业信息化建设的进程。

3.2 软件系统结构

软件体系结构是一个程序或系统构件的组织结构。通俗地说,一个软件体系结构是由一组构件、连接件和它们的约束组成的,同时还包括系统需求和结构元素之间的对应关系。软件体系结构在软件的整个生命周期中,是需求分析和软件分析、设计的纽带,对整个软件开发的成败起到关键性作用,因此软件体系结构是软件工程的核心,如何设计一个高效可靠的体系结构对于软件开发工作必不可少。

社区矫正是司法行政主要工作之一,其管理信息系统由国家司法部到省、市、县及乡镇级机构协同建立和管理监督。因此,需要具备良好的可扩展性和维护性。为了满足这个需求,社区矫正系统软件体系结构设计为分层结构,以便多级、多个子系统集成和软件维护;同时在表示层留有接口,方便与其他辅助系统进行连接。

社区矫正管理信息系统的软件体系结构分为表示层、应用层、Web 服务层、业务逻辑层、持久层、数据层以及基础架构层,如图 3-2 所示。通过基础架构层提供硬件、软件和网络支持,将数据传输到数据层;通过持久层的建模,业务逻辑层的配置和转换,由 Web Service 服务封装,应用层调用不同的服务,在表示层显示出来。

基础架构层:包括硬件、软件以及网络。硬件包含各种服务器、处理器等配套设备。软件包含操作系统、服务器应用程序等软件。网络层包括 Internet 接口、无线移动网络和司法专用网络,为系统提供网络支撑。

数据层:系统数据库采用 MySQL 数据库,为系统提供数据的存储、维护以及安全保障。数据层是矫正工作的数据核心,存储矫正对象的所有基本信息和矫正资料,为矫正工作提供数据支持。

持久层:采用 Hibernate 架构实现持久层的数据库映射,对数据库中的字段进行封装,解决数据处理难的问题。

业务逻辑层:通过 Spring 架构管理系统的业务逻辑,组件协作对象以及 DAO 组件,调用持久层的类模型,为上层提供接口和业务逻辑。

Web 服务层:对系统的业务进行服务封装,采用 Web Service 技术,将系统功能封装成服务,为上层应用层提供服务支持。

应用层:包含具体的应用,包括位置监控、警示告知、信息管理、档案管理以及系统管理等。

表示层:提供系统可视化显示,以 JSP 和 ExtJS 架构为主,为用户提供了美观、人性化的可操作界面。

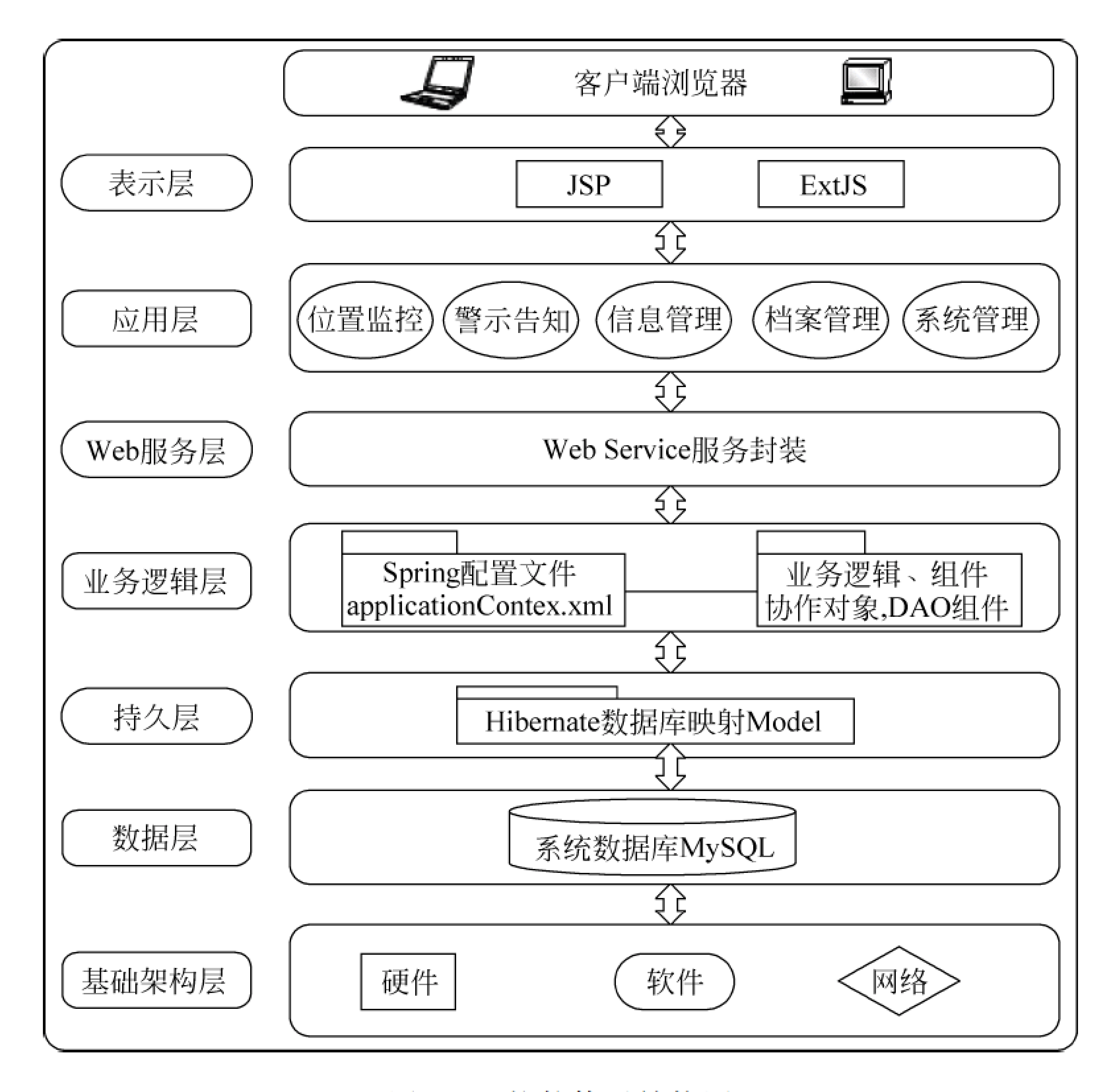


图 3-2　软件体系结构图

3.3　网络支撑结构

3.3.1　网络体系结构

网络体系结构是指计算机网络系统的整体设计，它为网络硬件、软件、协议、存取控制和拓扑提供参考模型。社区矫正工作要求各个省、自治区、直辖市所辖各地(市、州)建立自己的社区矫正网络，并连接到上级网络系统，形成全国性的社区矫正网络系统。因此，网络体系结构包括司法政务外网、政府公务网接口、互联网接口等。

社区矫正网络的体系结构，如图 3-3 所示。

在如图 3-3 所示的社区矫正网络体系结构中，政务外网与应用服务器之间交互设有防火墙、传输服务器，通过这些网络设备来满足网络安全与传输需求。通过司法政务网中的应用服务器来管理矫正信息系统的数据，通过数据传输系统进行与政府公务网之间的数据交换，并备份数据到中心服务器上。司法政务网和政府

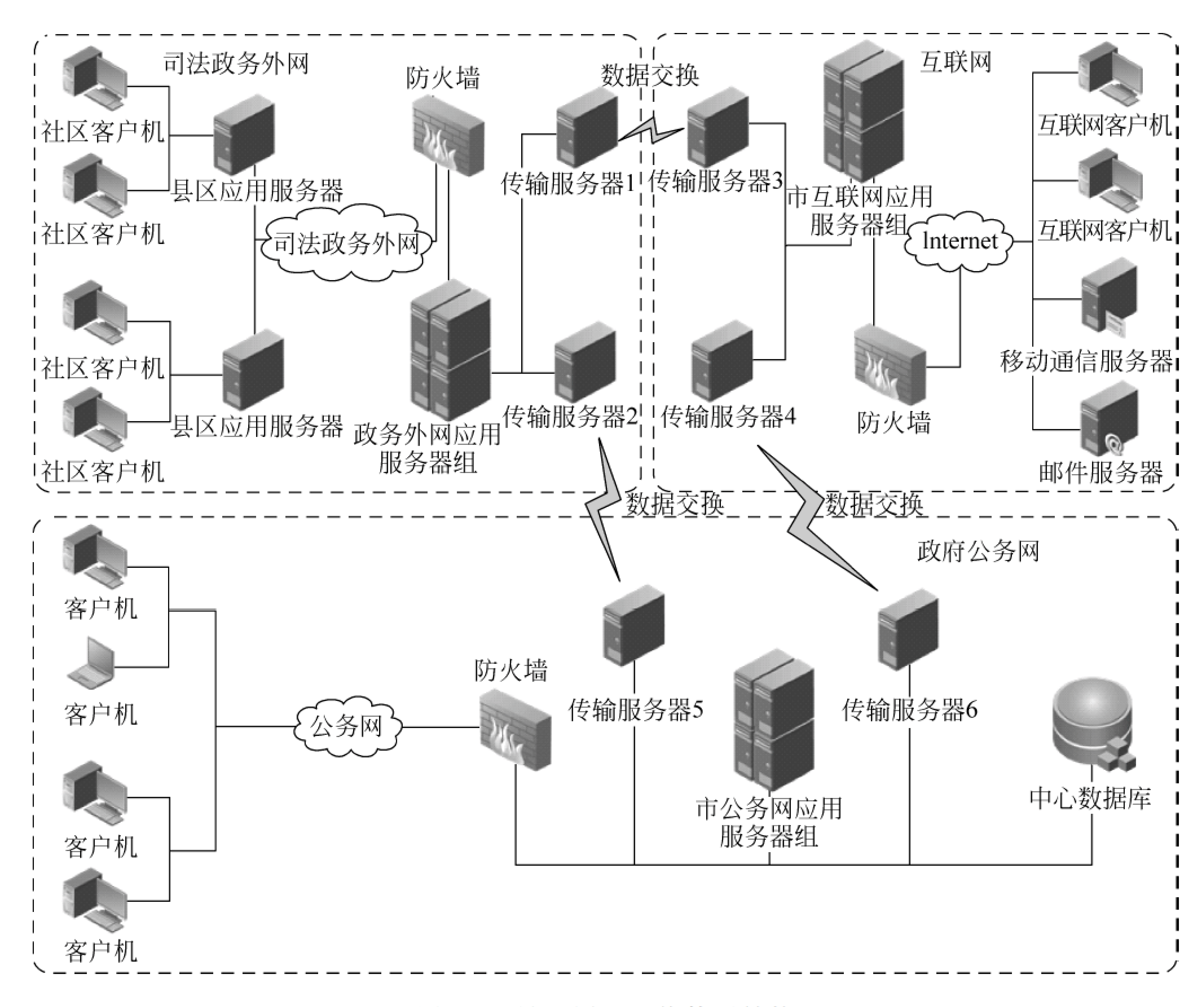


图 3-3　社区矫正网络体系结构

公开网都可以通过互联网与社区矫正网络互连互通。

下面介绍县(市、区)辖区内的社区矫正子系统网络结构,如图 3-4 所示。

如图 3-4 所示的社区网络是小型网络,包括数据库服务器、邮件服务器、应用服务器、移动网络以及社区矫正管理信息系统软件。系统将矫正对象的各类数据实时传回至系统管理平台,由系统服务器完成数据存储和处理。按照要求,社区矫正对象配备具有 GPS 定位功能智能手机和电子手环。各司法所社区矫正管理人员通过监管平台对社区矫正人员实行区域监管定位,通过二维地图实时了解社区服刑人员所在位置及一段时间内的移动轨迹,做到全方位的监管。当社区矫正对象超出所设定的区域时,系统会自动警告越界者,并通知矫正管理人员,同时在电子地图上自动显示越界时间、地点。社区矫正管理信息系统还为每一位社区矫正对象建立电子档案、矫正文书、日常管理等多项信息档案,为实现社区矫正管理提供基础数据。

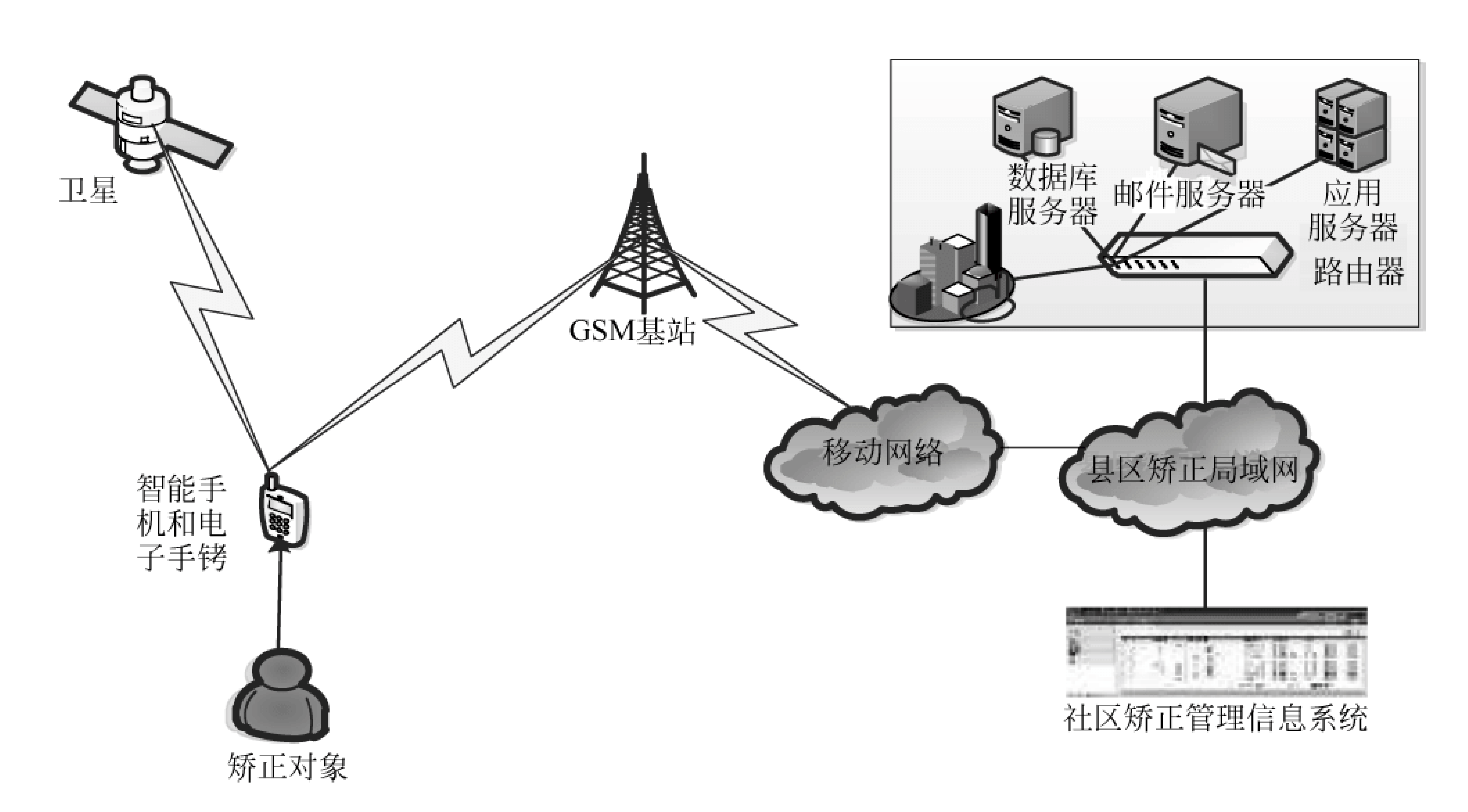


图 3-4　社区网络体系结构图

3.3.2　硬件体系结构

硬件结构是社区矫正中位置监控系统的基础，为位置监控系统提供底层支撑和数据收集设施，对整个系统来说至关重要。硬件体系结构支持实现监控，获取矫正对象位置信息，并传送给社区矫正位置监控系统。系统对收集的数据数据进行处理，并展示给矫正管理员，为矫正监控提供依据。社区级矫正信息系统硬件体系结构如图 3-5 所示。

图 3-5 中的硬件设备包括电子手环、阅读器、定位器等，这些设备协同工作，组成社区矫正监控系统，共同完成识别定位功能。阅读器通过天线识别矫正对象的电子手环，定位其位置信息，并将该信息反馈给社区矫正位置监控系统。在后台系统中，管理员会实时查看矫正对象的位置信息和活动信息等历史记录。

读写器通过天线接收信号，将能量和数据传递给电子标签的天线。电子标签获取其距离等位置数据，并通过天线反馈给读写器，最后读写器通过接口与社区矫正位置监控系统进行数据交互。定位器的主要功能是辅助读写器获取更准确的位置信息。此外，在读写器范围之外通过手机获取位置信息，将这些信息最终传递到移动平台的服务器，包括地图服务器、短信服务器以及移动通信服务器，服务器上所有数据通过位置监控中心软件进行管理，通过终端反馈给管理员。

系统使用具有射频定位功能的电子手环作为电子标签，由于其射频识别的范围有限，并且过多阅读器导致成本增加。因此，系统除了电子手环外，还设计有特制的智能手机，在该智能手机上配置有专门的辅助定位软件，在射频识别不到的地

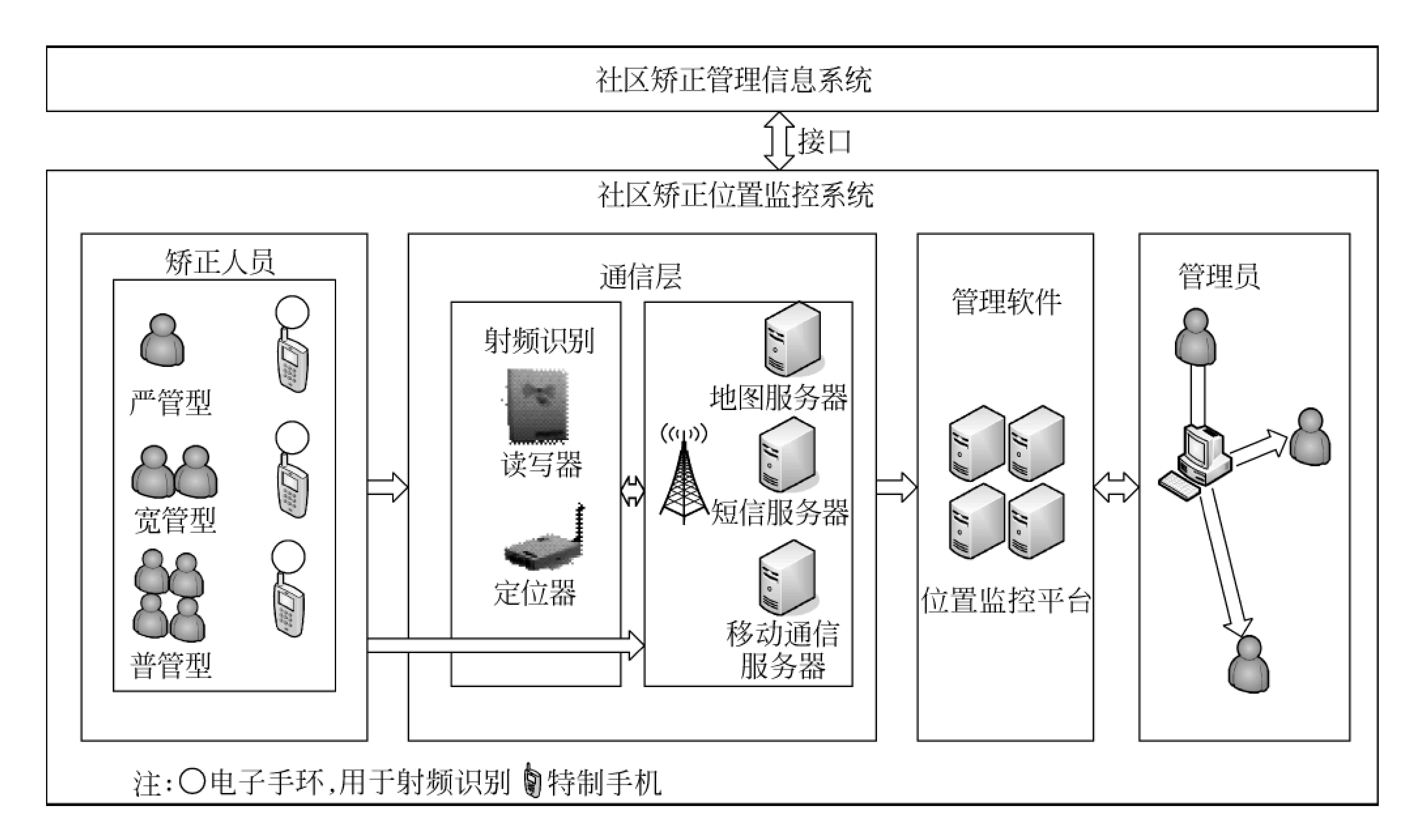


图 3-5　硬件体系结构图

方可自启动该软件获取位置信息,并传送到后台服务器。该软件还有内置数据库,将每个时间段的位置信息保存起来,当传送失败会在空闲时间重新传送。同时,智能手机和电子手环之间还有身份的确认,确保该手机是其本人携带。

3.4　关键技术

在社区矫正管理平台中,涉及矫正对象的定位技术及平台开发采用的 SOA 技术等。

3.4.1　无线定位方法

目前的定位技术按照是否测距来划分,主要可以分为基于测距算法(Range-based)定位和无须测距算法(Range-free)定位两大类。基于测距的算法通过测量节点间的距离或角度信息,使用三边测量、三角测量或最大似然估计等定位算法估计节点位置。常用的测距技术有 TOA 、TDOA、AOA 和 RSSI 等。无须测距定位算法则不需要距离和角度信息,算法根据网络连通性、链路质量或场景分析等特征信息来实现节点定位。

1. TOA 定位方法

TOA(Time of Arrival)定位方法,也被称作 TOF(Time of Flight),是一种通

过计算信号由发射器到达接收点的延迟时间来确定二者相对距离的方法。由于一般情况下超声波、电磁波等在空气中的传播速度是确定的，因此，信号发射点到接收点的相对距离可以通过物理学中的时间-速度公式求得：$d=vt$。在理想状况下，通过时延计算的相对距离能够满足定位的需要。

当采用一个接收器时，目标可能出现的位置在以接收器为中心、以二者相对距离为半径的圆形轨迹上，因此无法确定目标的确切坐标。当采用两个接收器时，两个接收器形成的圆形轨迹有可能相交于两个点，因而目标的位置也无法进行准确的判定。通过理论和实践可知，对于一个待定位的目标来说，至少需要3个不在同一直线的接收器就可以实现精确的定位任务。假设待定位的目标为O，3个接收器分别为R_1、R_2、R_3，那么采用TOA定位的原理如图3-6所示。

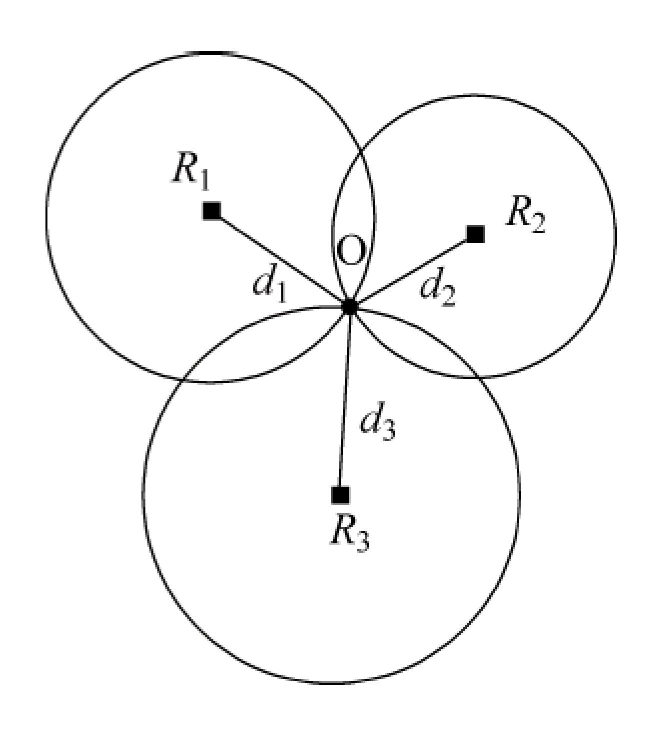


图 3-6　TOA 定位原理

其中，d_1、d_2、d_3分别为各个接收器与目标O之间通过信号的时间延迟计算的相对距离，每个接收器以相对距离为半径形成一个圆形轨迹。假设接收器R_i的坐标为(X_{ri},Y_{ri})，定位目标的估计坐标为(X_o,Y_o)，由于每个接收器在定位空间部署时位置确定，因此R_i的具体坐标在计算过程中为已知。利用三个圆形方程能够计算出唯一的交点，计算公式如下：

$$\begin{cases}(X_o-X_{r1})^2+(Y_o-Y_{r1})^2=d_1^2\\(X_o-X_{r2})^2+(Y_o-Y_{r2})^2=d_2^2\\(X_o-X_{r3})^2+(Y_o-Y_{r3})^2=d_3^2\end{cases}\tag{3.1}$$

从以上的描述和计算方法可以看出，使用TOA进行定位相对来说比较简单，在某些超声波定位中，由于声波波速较小，基本可以满足对物体定位的精度要求。例如在Active Bat Location System中就采用了超声波的TOA方式来对物体进行定位。

对TOA定位的另一个典型应用是GPS(Global Positioning System)系统。该定位系统由27颗卫星组成。只要拥有一个简单的手持设备，便可以对全球任何位置进行定位。由于是进行空间定位，即进行三维定位，而且需要补偿地球自转带来

的误差,所以至少需要接收四颗卫星的信号才可以进行较精确的定位。

但上述方法实际应用中存在两个问题,影响了方法的使用性。

(1) 时钟精度:因为无线信号的传播速度很快,又考虑到各种延迟,所以为了减小测量误差必须使用高精度的时钟,时间单位采用 ns,这对时钟硬件的要求过高,很难广泛的被移动设备所采用。

(2) 时钟同步:参与同一个定位过程的参考点之间必须保证时钟的同步,这样才能保证测量结果的正确性和精度。

在实际应用中,如果信号传播速度较慢或定位精度不高的远距离定位时,采用 TOA 方法进行定位是一种较好的定位方法。

2. TDOA 定位方法

TDOA(Time Difference of Arrival)定位方式,最早被雷达系统所采用的定位方式。与 TOA 类似,TDOA 定位所使用的观测值也是发射点到接收点的时间延迟。然而不同的是,TDOA 在定位过程中需要的是两个接收点的时间差值。

设两个基站 BS_1 和 BS_2 与移动基站 MS(Mobile Station)的距离分别为 R_1 和 R_2,如果已知距离之差:$R_{21}=R_2-R_1$,那么根据双曲线的定义可知移动台一定处在以两个基站作为焦点、与两焦点之间的距离之差恒定等于 R_{21} 的双曲线对上,如图 3-7 所示。

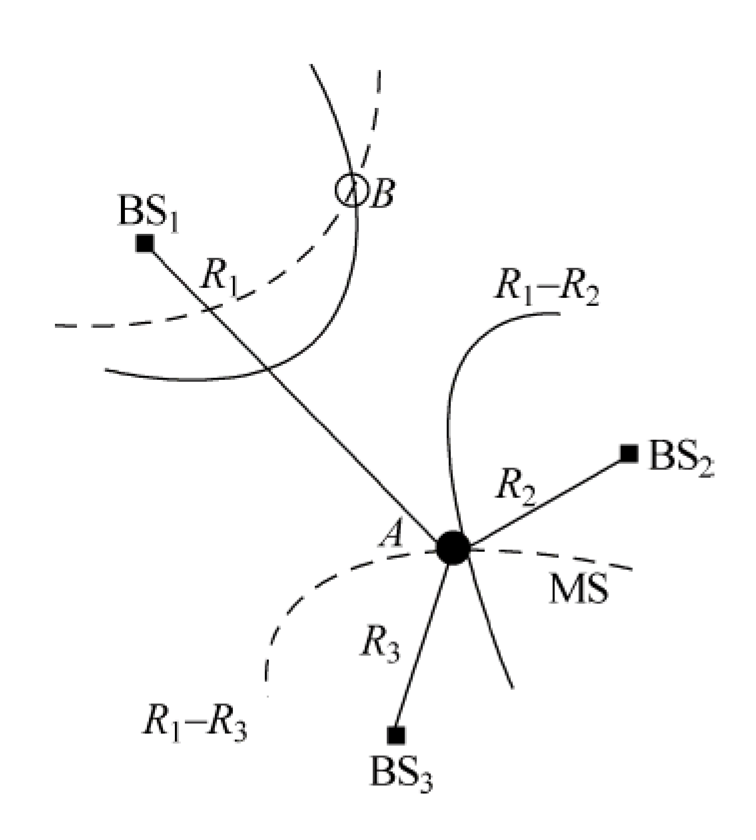


图 3-7 双曲线定位法

类似与 TOA 的定位方法,基站 BS_1 和基站 BS_2 与 MS 的距离之差可利用测量获得,也就是测量两基站发射信号到达 MS 的时间之差 t_{21},或是 MS 发射到两基站的时间差 t_{12}。显然,$R_{21}=c\times t_{21}$,c 为电波传播速度,值为 3×10^8m/s。双曲线定位方法中 MS 的坐标(x,y)与 BS_i 的坐标$(x_i,y_i)(i=1,2,3)$关系有式(3.2):

$$\begin{aligned}\sqrt{(x-x_2)^2+(y-y_2)^2}-\sqrt{(x-x_1)^2+(y-y_1)^2}&=R_{21}\\ \sqrt{(x-x_3)^2+(y-y_3)^2}-\sqrt{(x-x_1)^2+(y-y_1)^2}&=R_{31}\end{aligned} \tag{3.2}$$

求解以上方程组能够获得两解，各自相对应图 3-7 所标出来的两交点 A 和 B。而两解中只有一个为真实位置，这就需要提供某些先验信息（比如小区半径）来判断出真实的解，从而消除位置的模糊性。一些文献也将双曲线定位法称为基于电波到达时间差（TDOA）的定位法，即 TDOA 定位法，该方法是一种在蜂窝网中被广泛采用的方法。

3. AOA 定位方法

AOA（Angle of Arrival）定位方法，通过测定天线阵列中无线电波传播时到达的方向来进行测距的一种方法。AOA 方法通过天线阵列中的每一个天线接收信号的时间差（以 TDOA 的方式）来判定信号的方向。

这一方法是通过测量移动台发射电波到基站接收机天线或者天线阵列的入射角度，构成移动台到接收机之间的连线，这样就形成了方位线，可以定位方位角。如果可以获得两个或者更多个接收机天线提供的方位角测量值，即可以求出交点位置，从而确定移动台的待估计位置，如图 3-8 所示。

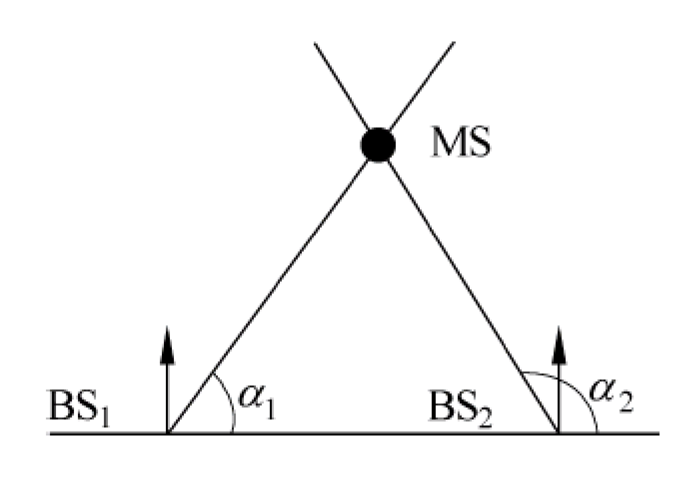


图 3-8　方位测量定位法

假设有两个基站：基站 AS_1 和基站 AS_2，并且分别测得移动台发出电波信号的到达角度分别为 α_1 和 α_2，则式(3.3)成立。

$$\tan(\alpha_i) = \frac{x - x_i}{y - y_i}, \quad i = 1,2 \tag{3.3}$$

通过求解上述非线性方程，可以得到移动台位置(x,y)。

AOA 定位的环境部署比较简单，使用两个接收器就可以实现对目标的定位工作。当目标发射无线信号时，接收器能够通过天线计算出目标所处的方位和相对夹角 α_1 和 α_2，从而通过两个夹角方向的交汇得到最终的位置坐标。因此，可以看出与 TDOA 等技术的定位体制相比，AOA 系统结构简单，但对硬件设备的要求很高，而且当定位环境比较复杂时，尤其是在城市建筑物比较密集的地区，存在比较严重的多径效应，使得定位的误差比较大。也正是由于上述原因使 AOA 系统在室内定位中尚未得到较大规模的普及。

4. 基于 RSSI 的测距定位方法

RSSI（Received Signal Strength Indication），即接收信号强度指示，具体指（前

向或者反向）接收机接收到信道带宽上的宽带接收功率，无线发送层的可选部分，用来判定链接质量，以及是否增大广播发送强度。无线通信设备可以比较容易的获得信号强度，无须购置额外设备。但是，信号强度受到多径效应的影响，使用相同频段的设备的信号之间会有相互干扰，使得信号强度波动明显，受环境因素影响较大。

1）对数路径损耗模型

对数路径损耗模型（Log-distance Path Loss Model）是一种用来预测信号在室内或者稠密人群环境下沿着某特定路径下随距离增加平均衰减程度的传播模型。通过对信号传播模型的理论分析和实验测量，结果显示无论在室内还是在室外环境下，平均信号强度随着距离的增加以对数方式衰减。对于任意的 T-R（发送-接收站）距离，大尺度路径平均损耗可以用式（3.4）来表示：

$$\overline{\mathrm{PL}}(d)_{\mathrm{dB}} = \overline{\mathrm{PL}}(d_0) + 10n\lg\frac{d}{d_0} \tag{3.4}$$

其中，n 是路径损耗指数，它表示随着距离增加路径损耗变化速率。d_0 是近地参考距离，一般为 1m。d 是 T-R 间隔距离。横杠表示给定值 d 的所有可能路径损耗的总体均值。当按照对数-对数为比例绘制坐标时，该模型路径损耗是一条斜率为每 10 米 $10n$ dB 下降的直线。n 的值依赖于特定的传播环境。例如，在自由空间下，$n=2$；当遇到障碍物时，n 值会变大。

对数路径损耗模型只是针对实际空间中表示任意距离的平均信号损耗。通过上述公式只能对平均信号损耗进行预测。但是对于相同的 T-R 间隔具体到某一次测量，测量到的信号值是剧烈波动的，这是由于传播空间中存在障碍物产生的阴影（Shadowing）和慢衰退（Slow fading）所致。研究者发现，对某一确定位置、任意的 T-R 距离 d，实际测量的路径损耗 PL（d）是随机的，且服从对数正态分布，可以使用式（3.5）进行描述：

$$\mathrm{PL}(d)_{\mathrm{dB}} = \overline{\mathrm{PL}}(d) + X_\sigma = \overline{\mathrm{PL}}(d_0) + 10n\lg\frac{d}{d_0} + X_\sigma \tag{3.5}$$

并且接收到的信号强度可以用如式（3.6）表示：

$$P_r(d)_{\mathrm{dB}} = P_{t\mathrm{dB}} + G_{t\mathrm{dB}} - \mathrm{PL}(d)_{\mathrm{dB}} \tag{3.6}$$

其中 P_t 是发射功率，G_t 是发射节点天线增益，X_σ 是平均值是 0、标准差为 σ 的服从正态分布的随机变量，单位为 dB。

现有的基于 RSSI 的定位系统存在着许多问题。由于各种特性所决定，电磁波在媒介中传播时面临着信号衰减、折射、反射、散射、绕射、多径干涉和多普勒频移等多种物理过程的影响。特别是障碍造成的多径干涉和阴影效应，造成信号强度波动时具有随机性的特征。在室内无线环境下，无线设备功率较小，覆盖面积较小，室内环境中由于墙壁、天花板、窗户及其他附属设施造成室内电磁波传播环境

非常复杂，大量人员随机的移动过程也会对电磁波的传播造成多径影响，同时大量同频率的信号之间存在着较强的信号干扰。这些因素造成了现有基于 RSSI 的定位系统定位精度较差。此外，现定位系统的规模相对较小，可扩展性较差。部分定位系统采用部署参考标签的方法来提高使用 RSSI 定位的精度，但是系统性能受到参考标签密度、部署布局的影响，定位效果参差不齐。

2）基于测距的 RSSI 定位方法

当定位系统性能参数确定之后，移动站点处的功率强度大小仅与其到基站的距离有关，则通过对移动站处 RSSI 值的测量，可以得到其与基站的位置关系，理论上可以利用至少三个基站来推定目标的位置。类似于 TOA 定位，基于 RSSI 定位的数学求解方程为式(3.7)：

$$\begin{cases}(X - X_{R1})^2 + (Y - Y_{R1})^2 = d_1^2 \\ (X - X_{R2})^2 + (Y - Y_{R2})^2 = d_2^2 \\ (X - X_{R3})^2 + (Y - Y_{R3})^2 = d_3^2\end{cases} \tag{3.7}$$

其中 d_1、d_2、d_3 分别表示移动站与基站 $R1$、$R2$、$R3$ 之间的距离。

基于 RSSI 的定位，是以电磁波强度值为观测量来实现的，因此容易受到 NLOS 等其他干扰因素的影响，但在视距传播较好条件下，且定位精度要求不高的应用环境中可以满足要求。

5. 基于场景分析的定位方法

1）概述

基于场景分析的定位方法一般是预先建立基于 RSS 参数值的先验分布图，然后根据概率模型、k-NN(k-nearest neighbor)模型等做出相应的分析，实现定位。换句话说，就是通过对定位环境进行形式化，用一些具体的、量化的参数描述定位环境，定位过程中根据待定位物体所在位置的特征进行匹配来确定物体的位置。基于场景分析的定位的核心是位置特征数据库和匹配规则，它本质上是一种模式识别方法。

基于场景分析的定位方法又称位置指纹法，可以被看作是让计算机先学习信号强度与位置间的内在规律，然后再推理的过程，这实质是一个模式识别的问题。

2）指纹定位原理

指纹定位技术是一种基于信号强度相似度匹配的定位方式，它一般分为两个阶段：离线数据采集和在线定位。

假设定位空间中有 N 个信号发射器，那么在空间中的每个采样点就可以接收到 N 个信号强度值。离线数据采集阶段的主要任务就是建立关于位置-信号强度关系的数据集，这些数据集就像指纹一样能够代表某一位置的信号特征。指纹定位的第二阶段主要目的就是根据定位目标接收的观测值在指纹数据集中查找最匹

配、最可能的位置-信号强度值,从而最终确定目标的坐标。基于指纹定位的原理如图 3-9 所示。

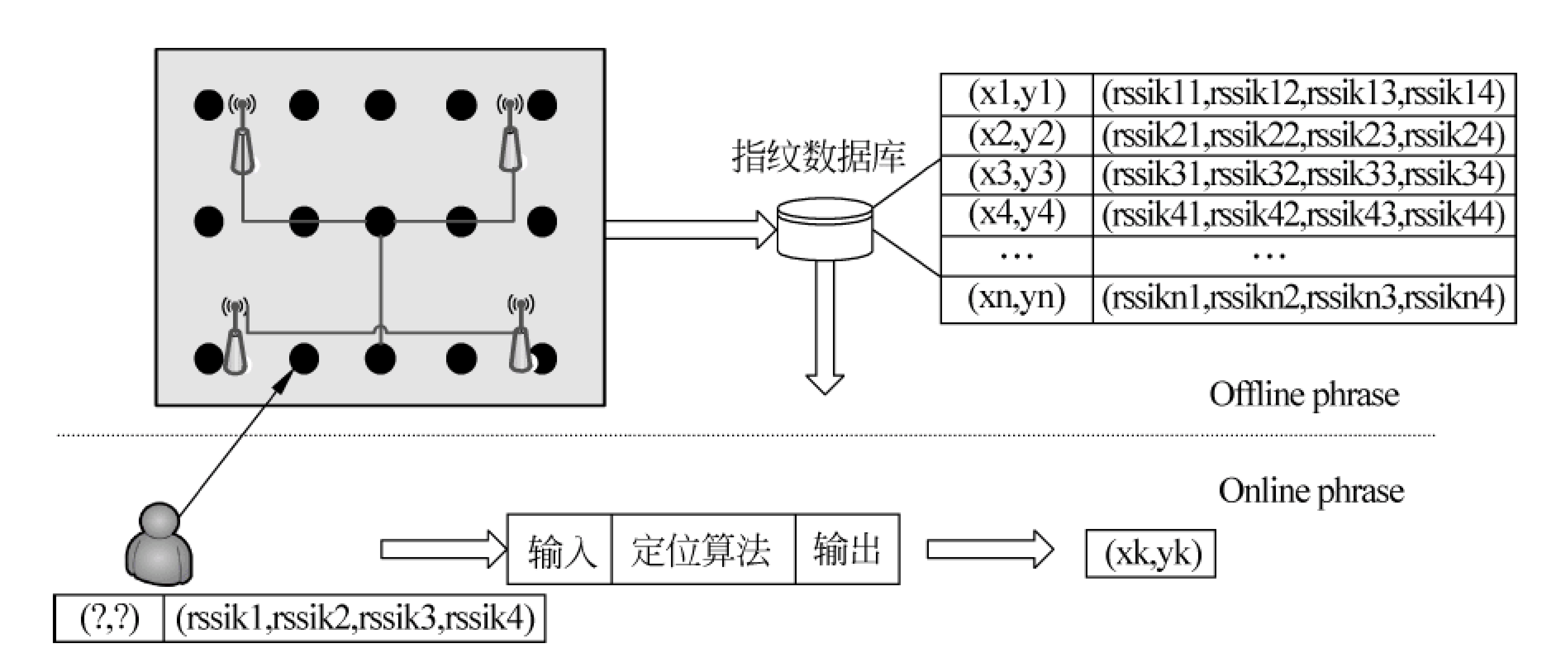


图 3-9 指纹定位原理

基于指纹的定位逐渐得到人们的认可,它能够解决信号传播模型定位的诸多缺点,能够处理室内环境比较复杂的情况,再加上一些新方法的引入,例如粒子滤波、图谱分解等,使得该方法的研究日益深入。但是指纹定位技术的缺点也很明显,最主要的问题集中在离线采样和信号的波动性上。一般来说,每个不同的定位环境都有一个指纹数据库,不同环境之间的数据基本上不能进行通用,这就造成了指纹定位较差的可扩展性。一旦环境改变,原来的数据就不可用,需要进行重新采集和更新,这种模式极大地增加了指纹定位技术在离线阶段的复杂性。除此之外,当同一个环境中需要采样的样本点非常多时,同样也会增加时间的复杂度。目前,对于指纹定位技术的研究主要集中在如何降低离线阶段的成本,以及如何利用观测数据进行较好的在线数据匹配上。

6. 被动定位方法

传统的定位方法中,定位对象需要携带一个具有无线收发能力的标识设备,例如传感器、有源标签以及其他智能设备。这些设备能够发射信号并能够被接收器所接收,系统利用信号强度可以实现位置的估计。但是,在很多应用场合中,例如紧急救援、安全监控、位置感知、入侵探测等领域,让目标携带设备以实现无线定位是不适用的。因此,如何实现无需目标配合的被动定位具有广泛应用前景。

被动定位就是移动对象不携带收发设备,它的行为轨迹被定位系统进行主动探测,它最早是由 Youssef M 提出的。被动定位方法主要利用移动对象对整个定位空间中信号强度分布的影响来确定移动对象的位置,通过分析对比信号强度变化的差异性,提取出信号强度变化区域的位置分布,继而能够得到较为准确的位置点。在这种方法中需要对整个空间的信号强度分布状态、影响因子以及空间部署

做出准确的判断。

当前，对于被动定位技术的研究取得了一些进展，国外相关领域的人员以及香港科技大学、中国科学院等研究机构的一些研究人员对此进行了初步的研究，形成了一些经验和方法。对于被动定位技术存在两种基本思路。一种是通过两个具有收发功能节点之间的链路状态变化来确定位置点。这种方法在一个区域中覆盖多个节点，两两组成一个链路。当有移动对象从链路附近经过时会产生链路之间信号的波动，从而确定出定位点出现在链路附近的区域中，如图 3-10 所示。

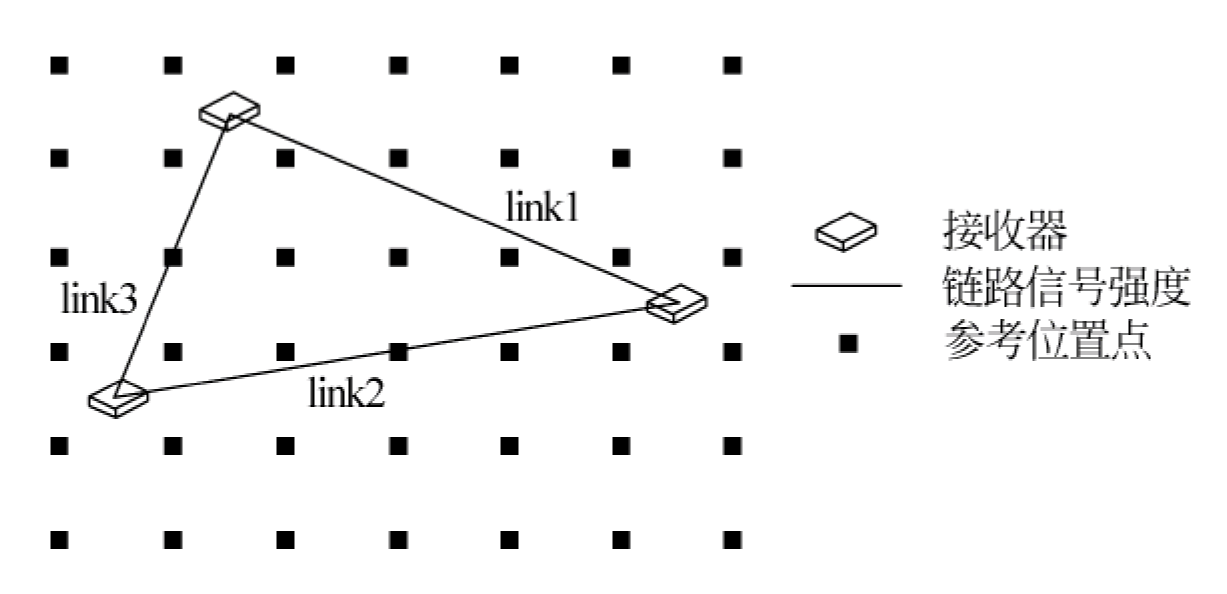


图 3-10　节点间链路信号波动定位

可以用两个节点 d 和 a 之间的信号强度差 $|RSS_d-RSS_a|$ 来表示链路间影响前后信号的变化值。然后设置衡量信号波动值的阈值(threshold)，并通过公式(3.8)对链路间的信号波动值进行检查：

$$\begin{cases} \text{yes} \ |RSS_d - RSS_a| > \text{threshold} \\ \text{no} \ |RSS_d - RSS_a| > \text{threshold} \end{cases} \tag{3.8}$$

当某一个链路的信号波动超过 threshold 时就可以判断对象可能出现在链路附近，于是开始进行定位过程。定位的实质就是建立链路信号波动与位置之间的数学模型，当得到一组链路波动值时，可以通过模型推导出具体的实际位置信息。

被动定位技术另一个比较可行的方法就是借助大量参考信标与接收器间的信号变化关系来定位，这种方式的主要出发点是判断哪些信标发生了信号变化。由于信标一般部署的较为密集并且这些信标的位置已知，因此如果能够确定出哪些信标接收信号发生了变化，那么就能够利用这些信标的位置来进行定位。前面的方法更加注重模型关系的建立，而这种方法更加关注信标的分类以及如何利用信标的位置来计算坐标。虽然这种方式依然依赖于收发设备间的信号变化，但是这种关系仅限于接收器去检测，如图 3-11 所示。

当移动对象在定位区域中移动时会引起读写器接收到的参考信标信号值发生变化，这些信标主要位于移动对象的轨迹附近。这样就可以将整个区域划分为两部分：不变的区域和变化的区域。在图中，区域 A 就是信号值不变的位置点。因此，实现定位的重点就是确定信号波动的边界线以及找出可能的位置点。这种定

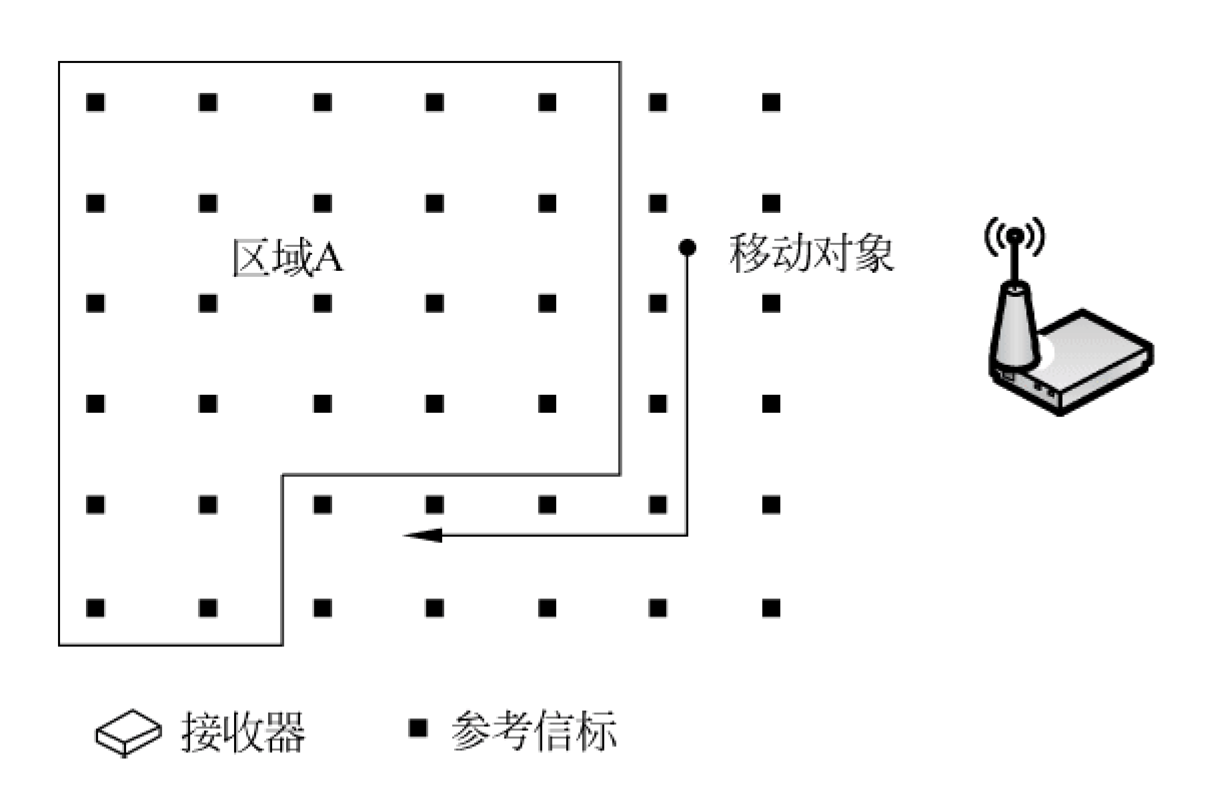


图 3-11　利用信标定位

位方法很适合对于移动轨迹的挖掘。

除了上述被动定位的基本思路之外，还有一些其他的定位方式被运，例如SPAN实验室的Patwari等建立的无线成像模型、kanso使用的压缩感知方法等，都对被动定位的发展和促进提供了多样化的解决方案。

7. 基于空间划分的定位

根据读写器自身参数的不同，如工作频率、增益系数等，读写器系统有其自身的识读范围，对该范围内的标签读写器可以正常地识读。根据这一特点，通过在定位空间中合理布置一定数量的读写器系统，用不同的读写器将定位空间划分成若干子区域，通过轮询所有读写器可以判定待定位标签所在的子区域。基于空间划分的定位可以满足一般低精度的定位需求，其系统原理如图3-12所示。

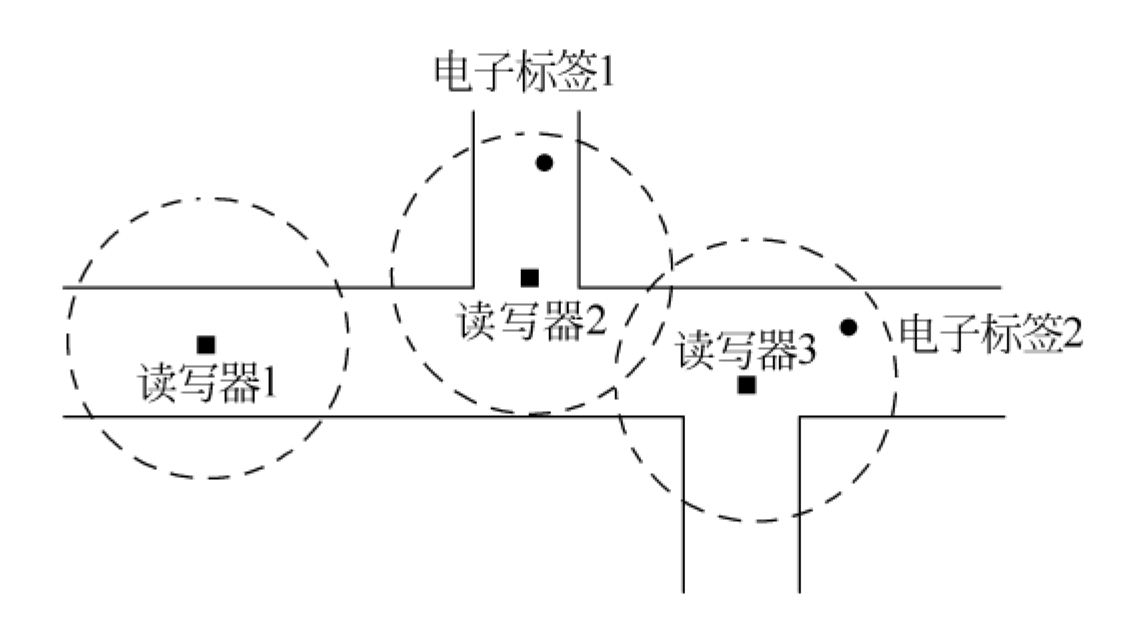


图 3-12　基于空间划分的定位原理

这种定位系统实施较为容易。由于需要大量读写器设备进行空间区域的划分，因此该系统实施成本较高。此外，该定位方法仅能体现定位目标的大致区域所在，因此定位精度一般，但对于某些精度要求不高的应用场合完全满足要求，如现在应用较为广泛的煤矿地下定位系统和楼宇定位考勤系统等。该定位方法实施的关键是合理部署阅读器硬件系统。

8. 常见定位技术比较

对现有的定位技术从设备的投入、定位精度以及对外界环境的干扰的适应程度等方面对比后可以发现，这些技术各有优劣。在不同的实际应用中，根据需求不同可以选用合适的定位技术，或者综合某几种技术来定制应用。

1）TOA（Time of Arrival）

需要涉及定位的节点之间能够精确的时间同步，这对时钟等设备精度要求非常高，购置设备费用较高。所以，针对电磁波设计的 TOA 系统投入实际应用有一定困难。比如，GPS 就是一个典型的应用，但其投入非常巨大。对于中短距离的定位应用中，产生的误差较大并且很难通过误差补偿的方式提高定位精度。

2）TDOA（Time Difference of Arrival）

受限于超声波传播距离有限（RTLS（Real Time Location System，实时定位系统）所使用的超声波信号通常传播距离仅为 10～15m，因而网络需要密集部署）和 NLOS（Non Line of Sight，非视距）问题对超声波信号传播的影响。

3）AOA（Angle of Arrival）

也会受到外界环境影响，而且需要额外硬件，对天线阵列的尺寸、数量和功耗有比较高的要求，需要天线对信号方向有较高的灵敏度。

4）RSSI（Received Signal Strength Indicator）

虽然符合低功率、低成本的要求，也无需额外的设备功能的支持，但由于信号由于多径、衰退、阴影等原因，存在着信号值的不确定性，受外界环境影响较大，并且信号间的干扰也是一个严重问题，这些因素有可能对采用 RSSI 的定位系统产生较大测距或定位误差。

表 3-1 分别从设备的投入、计算代价、定位精度、受环境影响程度和应用方面，对典型的 RFID 室内定位技术进行对比。可以发现这些技术各有优缺点，在实际应用中，根据需求不同可以选用合适的定位技术，或者综合某几种技术应用。

表 3-1　室内定位技术对比

定位技术	设备投入	计算代价	精度	受外界环境干扰	应用
AOA	需要额外硬件，对天线阵列的要求高，需要天线对信号方向有较高灵敏度	算法复杂度较高	在非视距环境下，误差较大	受外界噪声和 NLOS 非视距的影响	蜂窝网无线定位和军事领域，一般定位系统很难满足条件

续表

定位技术	设备投入	计算代价	精度	受外界环境干扰	应用
TOA	时钟等设备要求高，设备费用高。阅读器和标签有高精度的计时功能	算法实现容易，求解时间短	中短距离定位误差大	受传播路径的干扰大	GPS 应用
TDOA	需要网络密集部署，对系统计时精度要求高	需要大量计算和通信开销	较 TOA、RSSI 精度高，可达厘米级	受限于超声波传播距离有限和 NLOS 非视距问题对超声波信号的传播影响	Cricket 系统和 AHLos 系统
RSSI	无需额外设备投入，设备要求低，低功率，低成本，实用性高	RSSI 比较容易获得	适合中短距离定位，精度比较高	外界环境影响较大，信号多径、衰退以及信号间干扰	SpotON 系统 LANDMARC 系统和 VIRE 系统

3.4.2 全球定位系统

全球导航卫星系统(Global Navigation Satellite System，GNSS)是 20 世纪 90 年代中期国际民航组织以及欧洲空间局等倡导发展的一种全球性的位置和时间测定系统，主要包括全球定位系统(Global Positioning System，GPS)、全球导航卫星系统(GLONASS)、中国北斗卫星导航系统(BeiDou(COMPASS) Navigation Satellite System)、广域增强系统(Wide Area Augmentation System，WAAS)、EGNOS1 欧洲静地卫星导航系统、星载多普勒无线电定轨定位系统(Doppler Orbitograph and Radio Positioning Integrated by Satellite，DORIS)，以及正在建设的伽利略(Galileo)卫星导航定位系统和印度区域导航卫星系统(Indian Regional Navigational Satellite System，IRNSS)等。这些卫星系统全天候为全球陆海空各类载体(飞机、船舶、导弹、汽车及个人手持设备等)连续提供高精度三维位置、速度和精密时间信息。全球导航卫星系统是适用范围最广的三维空间定位技术。

GPS 由美国研制和维护，在轨卫星 24 颗，可以为地球表面 98%的地区提供准确的定位、测速和高精度的时间标准。移动目标最少只需接收到 3 颗 GPS 卫星信号就能迅速确定所处的位置。GPS 提供军、民两种定位精度，军码精度优于 10m，只供美军及其盟友使用；民码精度 20m 左右，已对全世界开放。GLONASS 由前苏联从 20 世纪 80 年代初开始研制，目前在轨卫星 12 颗。GLONASS 与 GPS 兼容，虽然精度低于 GPS，但 GLONASS 抗干扰能力强，并且打破了美国独家经营卫

星导航的局面，并且可与 GPS 结合获得更高的定位精度。GLONASS 已经于 2011 年 1 月 1 日在全球正式运行，目前有 24 颗卫星正常工作，其中 3 颗备用。

2000 年，我国发射了两颗北斗导航试验卫星，成为世界上第三个拥有自主卫星导航系统的国家。2004 年，我国正式启动了北斗卫星导航系统的建设。根据规划，2020 年前，该系统将提供覆盖全球的导航、授时和短报文通信服务。北斗卫星导航系统提供开放服务和授权服务两种方式。开放服务在服务区免费提供服务，定位精度为 10m。目前北斗卫星导航系统已在诸多领域发挥了重要作用。

1. GPS 系统

1957 年 10 月第一颗人造地球卫星上天，电子导航应运而生。美国从 1973 年开始筹建全球定位系统，1994 年投入使用。GPS 经历 20 年，耗资 300 亿美元，是继阿波罗登月计划和航天飞机计划之后的第三项庞大空间计划。

GPS 系统由以下三大部分组成。

1）宇宙空间部分

GPS 系统的宇宙空间部分由 24 颗工作卫星构成，最初的设计将 24 颗卫星均匀分布到 3 个轨道平面上，每个平面 8 颗卫星，但之后改为采用 6 轨道平面，每平面 4 颗卫星的设计。GPS 的卫星布局保证在地表绝大多数位置，任一时刻都有至少 6 颗卫星在视线之内，可以进行定位。

2）地面监控部分

GPS 系统的地面监控部分包括 1 个位于美国科罗拉多州 SChriever 空军基地的主控中心（Master Control Station，MCS）、4 个专用的地面天线以及 6 个专用的监视站。此外还有一个紧急状况下备用的主控中心，位于马里兰州盖茨堡。

3）用户设备部分

要使用 GPS 系统，用户端必须具备一个 GPS 专用接收机。接收机通常包括一个和卫星通信的专用天线，用于位置计算的处理器，以及一个高精度的时钟。随着技术的发展，GPS 接收机变得越来越小型和廉价，已经可以集成到大多数日用电子设备中，目前配备有 GPS 接收机的手机已不在少数。

当接收机捕获到跟踪的卫星信号后，就可测量出接收天线至卫星的伪距离和距离的变化率，解调出卫星轨道参数等数据。根据这些数据，接收机中的微处理计算机就可按定位计算方法进行定位计算，计算出用户所在地理位置的经纬度、高度、速度、时间等信息。

接收机硬件和机内软件以及 GPS 数据的后处理软件包构成完整的 GPS 用户设备。GPS 接收机的结构分为天线单元和接收单元两部分。目前各种类型的接受机体积越来越小，重量越来越轻，便于野外观测使用。

2. GPS 定位原理

GPS 定位的基本运作原理很简单，首先测得接收机与 3 个 GPS 卫星之间的距离，然后通过三点定位方式确定接收机的位置。实际的 GPS 系统中，根据参考卫星的空间坐标，以及到参考卫星的距离，可以在空间中确定出一个唯一的球面。3 颗卫星可以确定出 3 个球面，通常情况下，两个球面的交集是一个圆，3 个球面的交集是两个点。因为其中有一个点的位置在宇宙空间中——这显然不可能是接收机的位置，因此只需要选取靠近地面的那个点作为接收机的坐标即可。

每一颗 GPS 工作卫星都在不断地向外发送信息，每条信息中都包含有信息发出的时刻，以及卫星在该时刻的坐标。接收机会接收这些信息，同时根据自己的时钟记录下接收到信息的时刻。这样，用接收到信息的时刻减去信息发出的时刻，就得到信息在空间中传播所用的时间。将这个时间乘以信息传播的速度（信息通过电磁波传递，其速度为光速），就得到了接收机到信息发出时的卫星坐标之间的距离。

根据 GPS 的工作原理，可以看出时钟的精确度对定位的精度有着极大的影响。目前 GPS 工作卫星上搭载的是铯原子钟，精度极高，140 万年才会出现 1 秒的误差。然而，受限于成本，接收机上面的时钟不可能拥有和星载时钟同样的精度，而即使是微小的计时误差，乘以光速之后也会变得不容忽视。因此，尽管理论上 3 颗卫星就已足够进行定位，但是实际中 GPS 定位需要借助至少 4 颗卫星。换句话说，所处的位置必须至少能接收到 4 颗卫星的信号，方可以应用 GPS 来进行定位。这极大地制约了 GPS 的适用范围，当处于室内环境时，由于电磁遮蔽的效应，往往难以接收到 GPS 的信号，因此 GPS 这种定位方式主要在室外场景施展拳脚。其中最为典型的应用就是汽车导航。

综上所示，GPS 定位是以星站距离测量为基础的，利用了时间延迟求算距离的方法，即利用同步发出的信号达到接收信号地点的时间延迟推算距离（距离＝光速×时间延迟量）。

由无线电测距交汇的原理可知，利用 3 个以上地面已知点可交会出卫星位置，反之，利用 3 颗以已知空间位置的卫星，又可交会出用户接收机的位置。GPS 卫星定位系统便是利用这个基本原理进行空间定位。GPS 定位的基本原理是根据高速运动的卫星瞬间位置作为已知的起算数据，采用空间距离后方交会的方法，确定待测点的位置。

GPS 发射包含有卫星位置的测距信号和导航电文。如果 GPS 接收机在某一时刻可以同时接收 4 颗以上的卫星信号，就可以测出接收机位置到 3 颗及以上卫星的距离并解算出该时刻 GPS 卫星的位置坐标，这样就可以利用距离交会法计算出待测站的位置。

3.4.3 移动通信基站定位技术

虽然GPS得到了广泛使用,但是它并不能处理所有的情况。例如,在室内环境中,GPS的定位效果很差,甚至很多时候根本就无法进行GPS定位。此外,GPS接收机的启动也相对比较缓慢,往往需要3~5分钟的时间,因此定位速度也相对较慢。有时候,人们的应用其实并不需要GPS那么高的精度。另一方面,尽管随着GPS接收机的廉价化和小型化,配备GPS接收机的设备正变得越来越多,但始终还有大量的移动设备并没有集成GPS模块。移动通信设备对于定位有很大的需求,然而,并不是每一部手机都配备了GPS接收器。综合这两方面的因素,在很多时候人们都需要用蜂窝基站定位来作为GPS定位的补充。

蜂窝基站定位上要应用于移动通信中广泛采用的蜂窝网络,目前大部分的GSM、CDMA、3G等通信网络均采用了蜂窝网络架构。在移动通信网络中,通信区域被划分成一个个蜂窝小区,通常每个小区有一个对应的基站。以GSM网络为例,当移动设备要进行通信时,先连接所在蜂窝小区的基站,然后通过该基站接入GSM网络进行通信。换言之,在进行移动通信时,移动设备始终是和一个蜂窝基站联系起来的,蜂窝基站定位就是利用这些基站来定位移动设备的。

1. 移动通信网络简介

蜂窝网络是由提供移动台(通常是指手机)通信服务的众多部件和端口组成的,它集信息收发、控制、转换、路由、登记等功能为一体。

图3-13展示了蜂窝网络的体系结构,大致可分为三部分:用户随身携带移动台MS(Mobile Station)、基站系统BSS (Base Station System)和网络子系统NSS (Network Subsystem),另外,蜂窝网络与公共交换电话网络PSTN(Public Switched Telephone Network)等外部网络相连。

BSS又可细分为BTS(Base Transceiver Station)和BSC(Base Station Controller)。BTS负责移动台接入通信网络,一个BTS服务的覆盖范围称作一个小区(cell)。这样蜂窝网络的覆盖区域可以被划分为一系列的小区集合,称作位置区LAC(Location Areas)。BSC是控制一个或多个BTS的装置,负责分配信道、频率管理和移动台的信号监测。

蜂窝网络的核心功能部分是MSC(Mobile Services Switching Center),它有着路由、通话连接和信息收发的功能。一个MSC管理着多个BSC,同时又和其他的MSC以及记录器(Register)连接。蜂窝网络的位置管理通过HLR(Home Location Register)和VLR (Visitor Location Register)实现。HLR是永久性存储用户信息的大型数据库,这些信息中包括用户当前的位置信息。VLR是存储用户位置信息的动态链接库,记录和HLR类似的信息,但是只记录当前在其位置区内

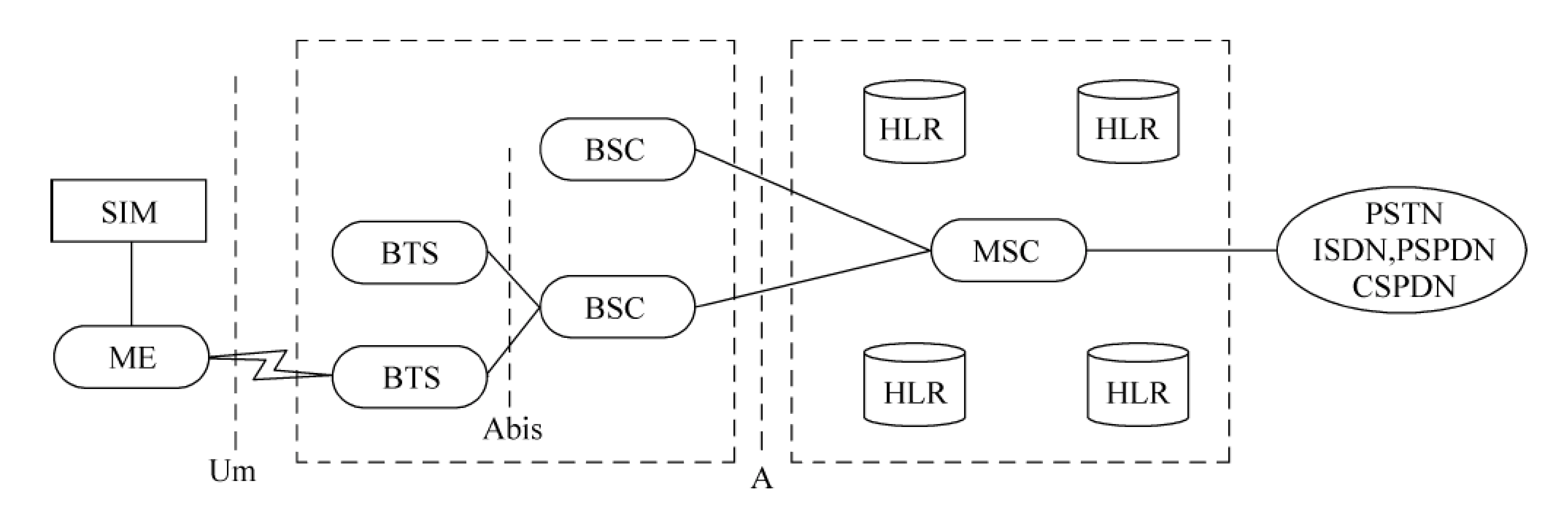


图 3-13 移动蜂窝网络体系结构

的用户的信息。

值得指出的是，借助手机和通信网络的信号交换，经定位处理可以在小区 Cell-ID 级别把手机定位在小区内。这是因为手机服务运营商（operator）记录了每个 BTS 的坐标，就可以确定每个与 BTS 连接手机的大致位置。

2. GSM 蜂窝基站的基础结构

GSM 网络的基础结构是由一系列的蜂窝基站构成的，这些蜂窝基站把整个通信区域划分成如图 3-14 所示的一个个蜂窝小区。这些小区半径小则几十米，大则几千米。

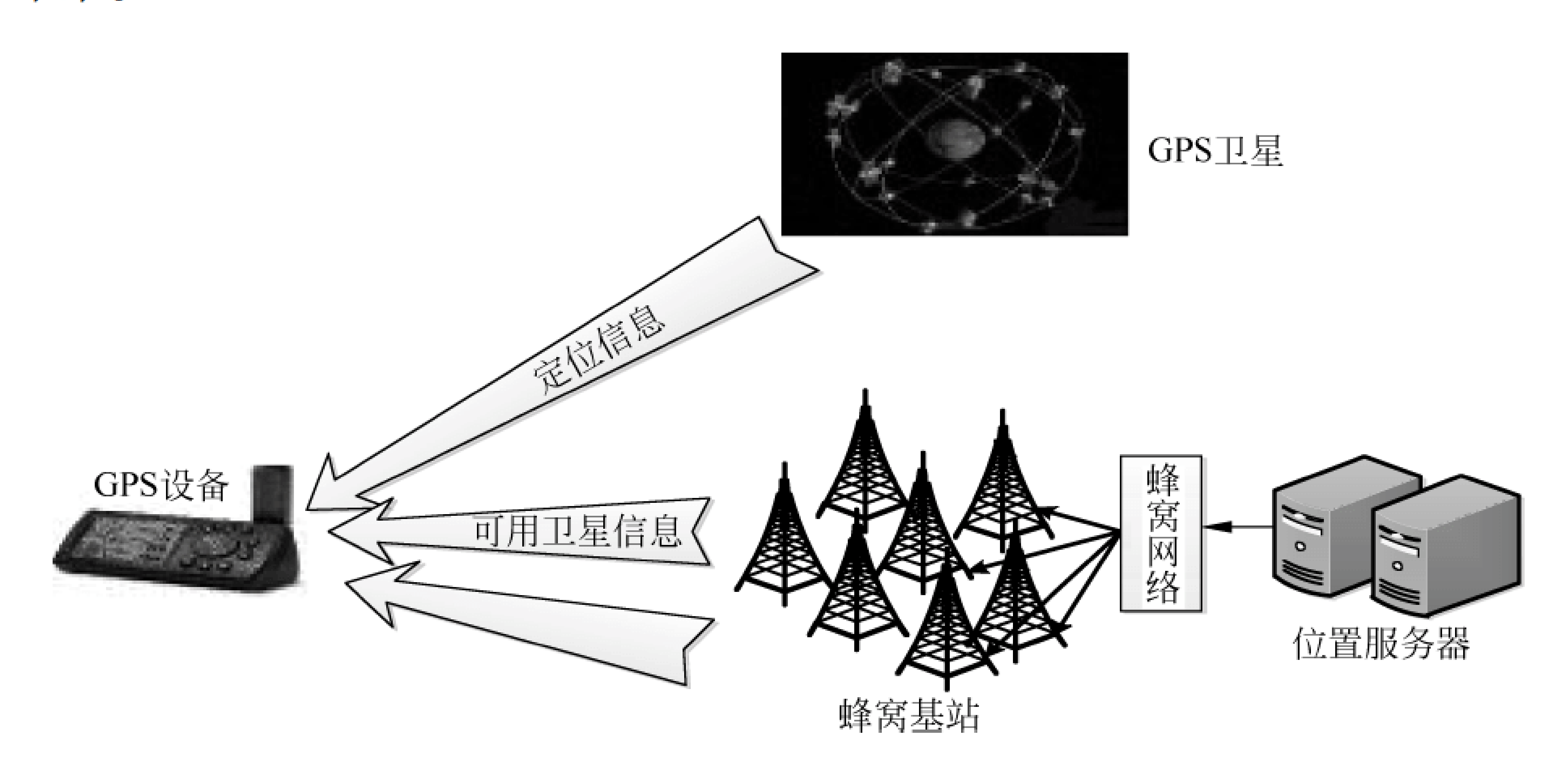


图 3-14 蜂窝基站

在 GSM 中通信时，总是需要和某一个蜂窝基站连接的，或者说是处于某一个蜂窝小区中的。那么 GSM 定位，就是借助这些蜂窝基站进行定位。

无线定位系统中对移动台的定位是通过检测移动台和多个固定位置的收发机之间传播信号的特征参数（如电波场强、传播时间或时间差、入射角等）来估计出目标移动台的几何位置。

3. 移动通信定位技术的基本原理

实际应用中,要确定目标的空间位置,要先测量某个与空间位置相关的物理量,然后应用一定的理论、计算方法或者数学模型,经过计算,最终得到目标的空间位置坐标。利用蜂窝移动通信系统对手机进行定位,就是通过测量表征通信基站和手机空间位置相关的物理量,再利用定位理论和数学模型计算出手机的空间位置。常用的定位技术有:基于场强的定位方法、基于起源蜂窝小区定位方法、基于到达角度定位方法、基于到达时间定位方法、基于到达时间差定位的定位方法。

1) 起源蜂窝小区定位

蜂窝小区技术(Cell of Origin,COO)起源于美国 E-911(移动位置服务),它是无线定位技术发展的第一阶段,也是这个业务平台采取的第一个定位方式。

COO 定位是一种单基站定位方法。这种方法非常原始,就是用移动设备所属基站的坐标视为移动设备的坐标。可想而知,这种定位方法的精度很低,其精度直接取决于基站覆盖的范围。如果基站荡盖范围半径为 50m,那么其误差最大就是 50m。在一些基站分布十分疏松的区域,一个基站覆盖范围的半径可达几千米,这个误差就相当大了。这种定位方法唯一的优势在于速度,通常只需要 2～3s 时间就可以完成定位,因此适用于情况紧急的场合。

COO 是蜂窝移动通信网络根据通信基站位置来定位手机的位置,其手机位置是在以其提供服务的通信基站为圆心、基站信号覆盖半径为半径的一个圆内,如图 3-15 所示。



图 3-15　COO 定位原理

从原理上可以看出,只要运营商支持,GSM 网络中的设备只要获取到当前接入基站的一个唯一代码,可以称之为基站 ID(或 CellID),然后根据这个 CellID 查出该蜂窝基站所在的具体地理坐标,即为手机当前的位置。起源蜂窝小区定位的最大优点是定位信令传输少,确定位置信息的响应时间快,一般在 3s 左右。另外,起源蜂窝小区不用对现有的手机和网络进行升级,只需在网络中增加简单的定位流程处理部分,就可以直接向现有用户提供位置服务。

起源小区定位方法 COO 在所有的蜂窝网络都适用。由于城区建筑密度大，普遍存在对通信信号严重遮挡和多路径干扰，某些具有高精度的定位方法(如 GPS)将失效，此时起源蜂窝小区定位方式将成为一种简捷、有效的定位方法，能够满足基本定位业务需要。

2) 增强观察时间差

增强观测时间差分(Enhanced Observed Time Different，E-OTD)技术的实现需要在大范围区域设置参考点，作为位置测量单元，每一个参考点都要有精确定时源，由于实际的蜂窝网络并不是完全同步的，该方案需要考虑时间同步的问题。当具有 E-OTD 功能的移动台和位置测量单元接收到来自至少 3 个基站信号时，从每个基站到达移动台和位置测量单元的时间差将被计算出来，这些差值可以被用来产生几组交叉双曲线，由此估计移动台的位置。E-OTD 方案的定位精度在 50～125m 之间，但是它的响应速度较慢，往往需要约 5s 的时间。这种方案需要对移动设备进行改进，这将意味着用户要付出较高的代价才能获得定位服务。因为这个原因，没有被 LBS 运营商大量使用。

3) TOA/TDOA 定位

基于网络的定位系统中通常采用精度较高的 TOA 或 TDOA 定位法。TOA 中，移动台位于以基站为圆心、移动台到基站的电波传播距离为半径的圆上。

在多个基站进行上述计算，则移动台的二维位置坐标可由 3 个圆的交点确定。TOA 要求接收信号的基站、移动台知道信号的开始传输时刻，并要求基站有非常精确的时钟。TOA 提供的定位精度比 COO 高，但是它的响应时间比 COO 或 E-OTD 更长(约 10s)。

TDOA 是通过检测信号到达两个基站的时间差，降低了时间同步要求。移动台定位于以两个基站为焦点的双曲线方程上，确定移动台的二维位置坐标需要建立两个以上双曲线方程，两条双曲线的交点即为移动台的二维位置坐标。在实际应用中通常采用最小均方误差算法，通过使非线性误差函数的平方和最小来估计移动台位置。特别是 TDOA 定位由于不要求移动台和基站之间的同步，在误差环境下性能相对优越，在蜂窝通信系统的定位中备受关注。

4) AOA 定位

基站通过阵列天线测出移动台来波信号的入射角，构成从基站到移动台的径向连线，两根连线的交点即为待定位移动台的位置，如图 3-16 所示。需要在每个小区基站上放置 4～12 组的天线阵，这些天线阵一起工作，从而确定移动台发送信号相对于基站的角度。当有多个基站都发现了该信号源时，它们分别从基站引出射线，这些射线的交点就是移动台的位置。AOA 的缺点是到达角估计会受到由多径和其他环境因素所引起的无线信号波阵面扭曲的影响，移动台距离基站较远时，

基站定位角度的微小偏差会导致定位距离的较大误差。

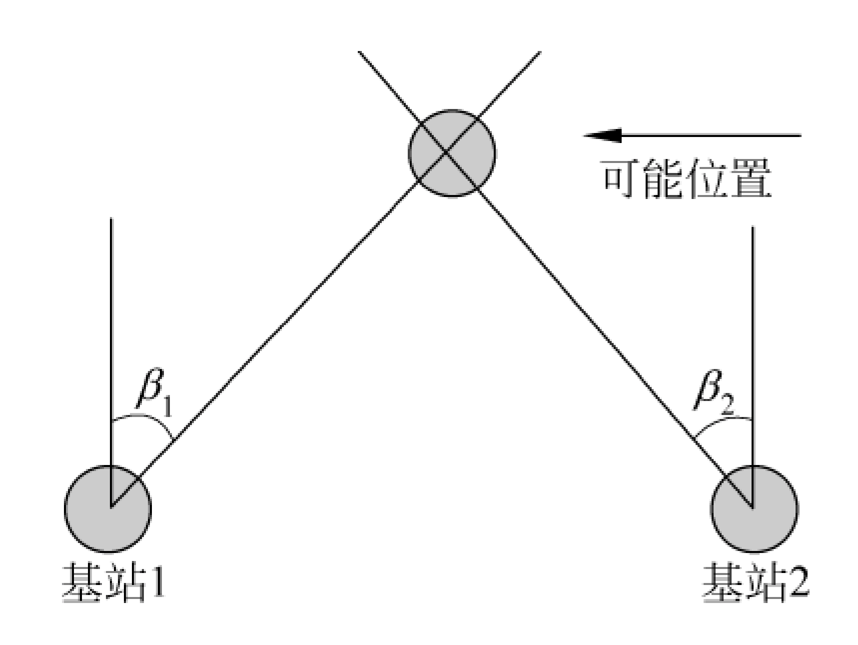


图 3-16　AOA 示意图

5）辅助 GPS

A-GPS 定位(辅助 GPS 定位)是一种混合定位，是 GPS 定位技术与 GSM 网络的结合，其定位系统结构如图 3-17 所示。A-GPS 具有很高的定位精度，目前正被越来越广泛地使用。

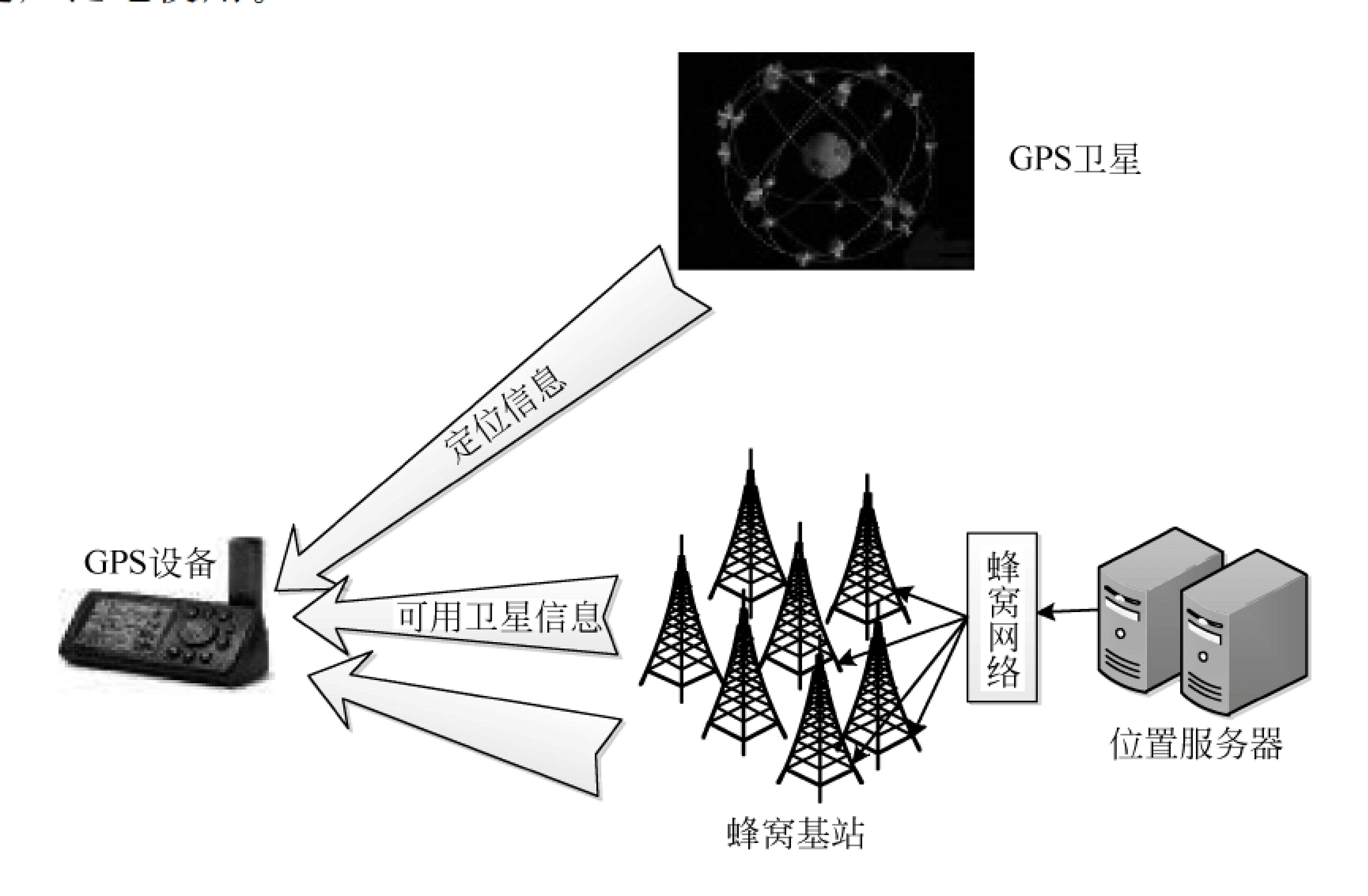


图 3-17　A-GPS 定位结构图

在 A-GPS 中，手机测量来自三颗或者更多卫星的信号到达时间。GSM 网络将 GPS 辅助信息发送到移动台，移动台得到 GPS 信息，计算出自身精确位置，并将信息发送到网络。A-GPS 有移动台辅助和移动台自主两种方式。

移动台辅助 GPS 定位是将传统 GPS 接收机的大部分功能转移到网络上实现。网络向移动台发送短的辅助信息，包括时间、卫星信号多普勒参数和码相位搜索窗口。这些信息经移动台 GPS 模块处理后产生辅助数据，网络处理器利用辅助数据估算出移动台的位置。

移动台自主 A-GPS 定位的移动台包含一个全功能的 GPS 接收器，具有移动台辅助 GPS 定位的所有功能，再加上卫星位置和移动台位置计算功能。

A-GPS 的优点是网络改动少，网络不需增加其他设备，网络投资少，定位精度高。由于采用了 GPS 系统，定位精度较高，理论上可达到 5～10m。缺点是现有移动台均不能实现 A-GPS 定位方式，需要更换，从而使移动台成本增加。

（1）A-GPS 定位基本机制。

硬件初始化（首次搜索卫星）时间较长、GPS 卫星信号穿透力弱及易受建筑物、树木等的阻挡而影响定位精度等缺点。A-GPS 定位技术通过网络的辅助，可以有效解决或缓解这些问题。对于辅助网络，以 GSM 蜂窝网络为例，一般是通过 GPRS 网络进行辅助。

（2）A-GPS 定位基本流程。

首先搜索卫星，A-GPS 定位仍然是基于 GPS 的，因此定位的首要步骤还是先搜索到当前地区的可用 GPS 卫星。

A-GPS 中从定位启动到 GPS 接收器找到可用卫星过程如图 3-18 所示，其基本流程如下：

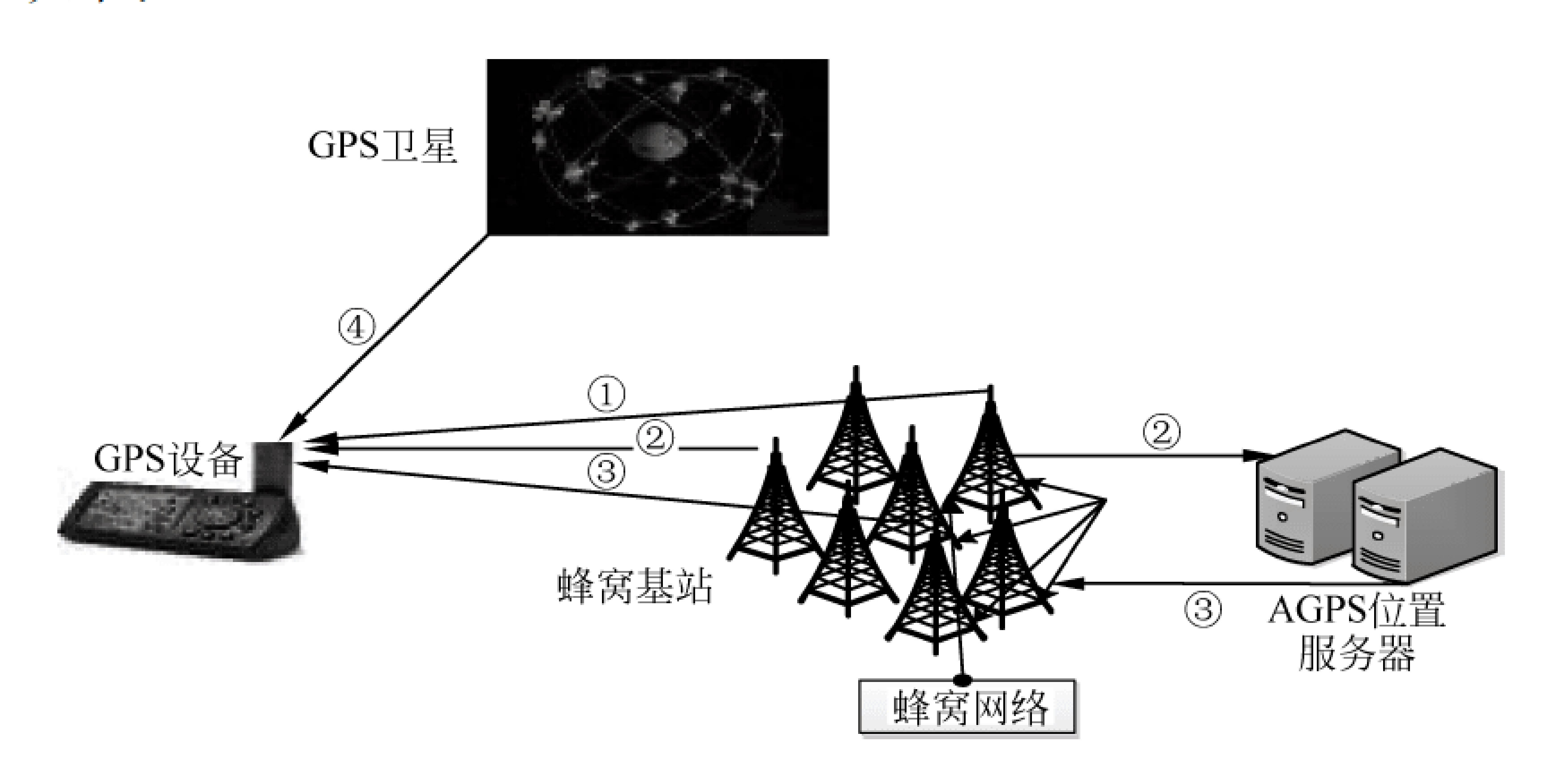


图 3-18　A-GPS 定位过程

① 设备从蜂窝基站获取到当前所在的小区位置（即 COO 定位）；

② 设备通过蜂窝网络将当前蜂窝小区位置传送给网络中的 A-GPS 位置服务器；

③ A-GPS 位置服务器根据当前小区位置查询该区域当前可用的卫星信息（包括卫星的频段、方位、仰角等相关信息），并返回给设备；

④ GPS 接收器根据得到的可用卫星信息，可以快速找到当前可用的 GPS 卫星；

至此，GPS 接收器已经可正常接收 GPS 信号，GPS 初始化过程结束。

然后计算位置，GPS 接收器一旦找到四颗以上的可用卫星，就可以接收卫星信号实现定位，计算过程如图 3-19 所示。

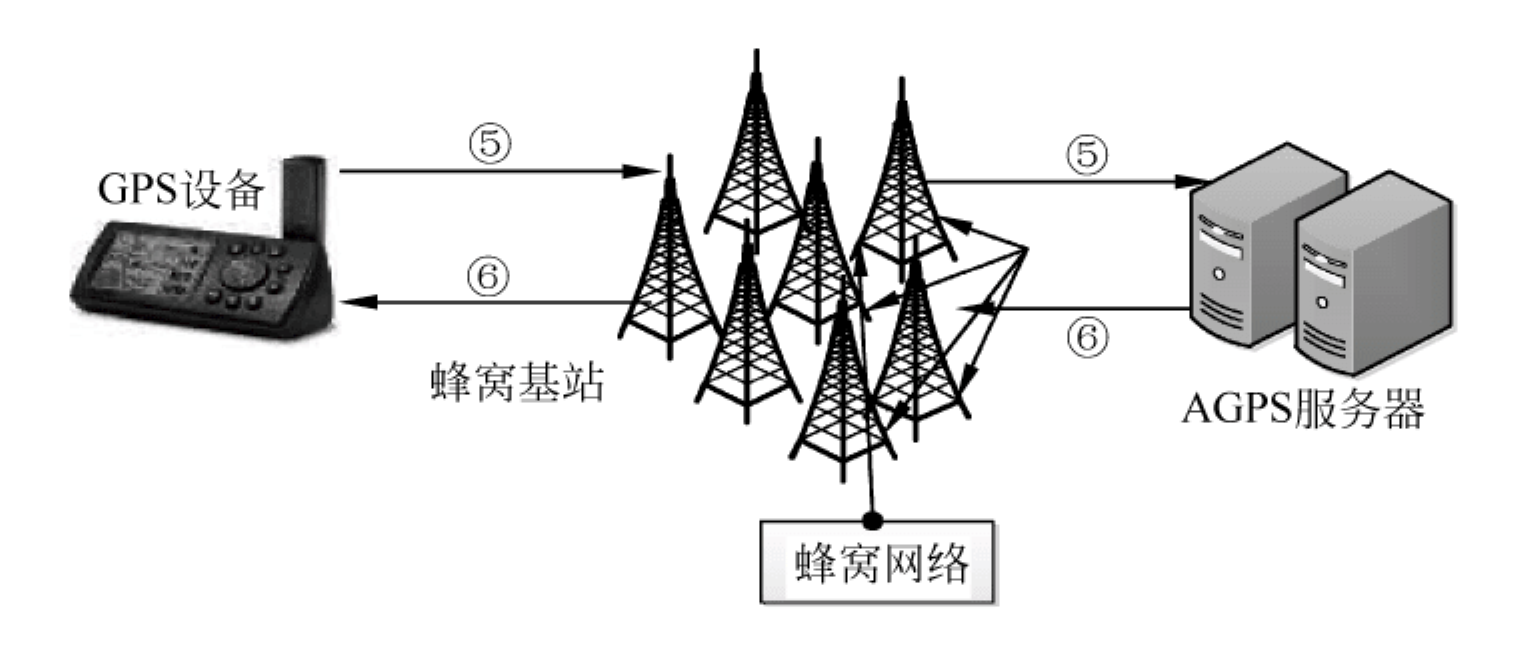


图 3-19　GPS 与 AGPS

接下来的过程根据位置计算所在端的不同，通常有两种方案：在移动设备端进行计算的 MS-Based 方式和在网络端进行计算的 MS-Assisted 方式。

⑤ 设备将处理后的 GPS 信息（伪距信息）通过蜂窝网络传输给 AGPS 位置服务器；

⑥ A-GPS 服务器根据伪距信息，并结合其他途径（蜂窝基站定位、参考 GPS 定位等）得到的辅助定位信息，计算出最终的位置坐标，返回给设备。

（3）A-GPS 的应用。

由于 A-GPS 需要网络支持，因此目前使用该技术的大部分设备为手机。目前大部分支持 A-GPS 的手机采用一种纯软件的 A-GPS 方案，该方案基于 MS-Based 位置计算方式。具体的方案为：定期下载星历数据到手机中，手机中的 A-GPS 软件会根据星历信息计算出当前位置的可用卫星信息，从而提供给设备用于快速搜星。

用户可以选择通过 Wi-Fi、固网等免费网络定期更新星历数据，避免使用蜂窝网络产生的数据流量费用。

3.4.4　SOA 技术

1. SOA 的概念

SOA（Service Oriented Architecture）即面向服务的体系结构，是一个组件模型。对 SOA 的理解多种多样，从技术角度看，SOA 就是一种体系架构，它描述了一种 IT 基础设施，使得不同的业务服务可以相互交换数据，参与业务流程，通过灵活的互相协作方式来完成具体的业务操作。这些业务服务独立于编程语言，独立于实现方法，独立于运行环境。

2. SOA 体系结构

采用 SOA 是为了解决社区矫正管理信息系统基于组件的分布式应用体系结

构所面临的问题。面向服务的体系结构基于"软件即服务"的思想,提出了一种新的解决软件重用和软件集成的方案。通过采用面向服务的体系结构,能够迅速便捷地构建开放的、模块化的、可重用的、与平台无关的、可扩展的分布式应用系统。作为 SOA 的一种实现手段,Web 服务提供了基于 XML 标准接口的若干中间件,具有完好的封装性、松散的耦合性、协议规范的标准性以及高度的可集成性等特点,能够很好地满足 SOA 应用模式的需求。

SOA 优点:代码重用,松耦合,平台独立,语言无关。

SOA 体系结构中共有三种角色,如图 3-20 所示。

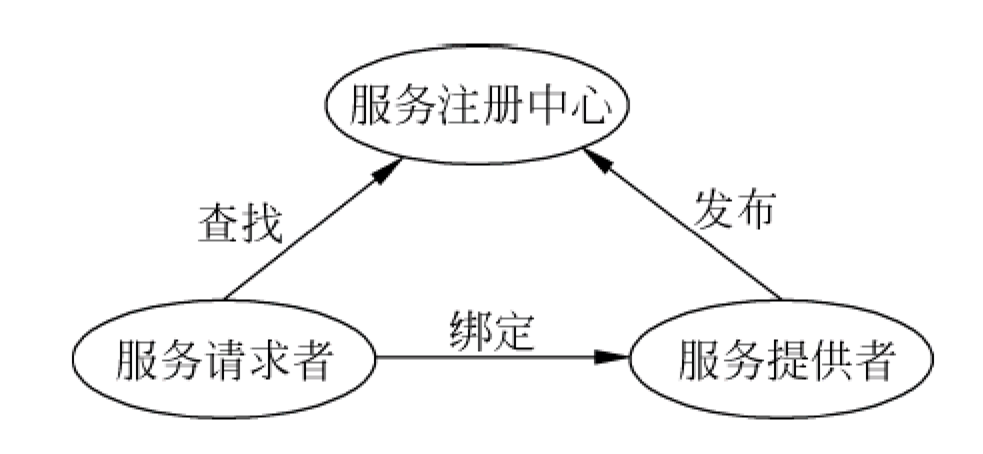


图 3-20 面向服务的体系结构

(1) 服务提供者:发布自己的服务,并对服务请求进行响应。

(2) 服务注册中心:注册已经发布的 Web Services,对其进行分类,并提供搜索服务。

(3) 服务请求者:利用服务注册中心查找所需要的服务,然后利用该服务。

这三种角色之间使用三种操作。

(1) 发布操作:使服务提供者可以向服务注册中心注册自己的功能和访问接口。

(2) 查找操作:使服务提供者可以通过服务注册中心查找特定种类的服务。

(3) 绑定服务:使服务请求者能真正使用服务提供者提供的服务。

3. Web Service 技术

Web Service 平台是一套标准,它定义了应用程序如何在 Web 上实现互操作性。Web Service 是技术规范,SOA 是设计原则。从本质上讲,SOA 是一种架构模式,而 Web Service 是利用一组标准实现的服务。Web Service 是实现 SOA 的方式之一。用 Web Service 实现 SOA 的好处是:可以实现通过一个中立平台来获取服务,以获取更好的通用性。

Web Service 核心技术包括:

(1) xml——Web Services 平台中表示数据的基本格式,它解决了数据表示的问题。

(2) SOAP(简单对象访问协议)——提供了标准的 RPC 方法来调用 Web

Services。SOAP 规范中定义了 SOAP 消息的格式，以及如何通过 HTTP 协议来使用 SOAP。SOAP 是基于 xml 语言和 XSD 标准的，其中 xml 是 SOAP 的数据编码格式。

（3）WSDL——一种基于 xml 的用于描述 Web Services 及其操作、参数和返回值的语言。

（4）UDDI——描述、发现和集成（Universal Description，Discovery and Integration）的英文缩写，它是由 IBM、微软等公司倡导的，其目的是在网上自动查找 Web Services。

Web Service 技术的主要目标是在现有的各种异构平台的基础之上构筑一个通用的平台无关、语言无关的技术层，各种应用依靠这个技术层来实施彼此的连接和集成。



第4章

社区矫正系统数据集成技术

4.1 社区矫正信息系统数据集成需求

数据集成是指把不同来源与结构的数据在逻辑上或物理上有机地集中起来，实现不同应用系统间的数据交换与共享，进行数据的同步化、标准化、转换、映射与传输。

由于社区矫正工作与监狱管理、安置帮教，以及法院、检察、公安等政府部门的业务工作紧密联系，需要在相关职能部门之间定期进行数据交换，以有效防止“脱漏管”。但是由于历史原因，不同时期开发的信息系统差异较大，存在多个不同软硬件平台的信息系统。这些系统的数据源彼此独立、相互封闭，数据难以在系统之间交流、共享和融合，形成了信息“孤岛”。尤其是司法行政、公安等政府部门、单位数据标准不统一、不规范的现象依然存在，信息“孤岛”还未消除。目前已有的社区矫正信息系统较难实现与监狱、公安等政府部门信息系统的集成与数据交换。因此，必须研究解决社区矫正信息系统与其他信息系统的数据集成问题。

社区矫正信息系统要求的数据集成内容主要包括调查评估信息、社区矫正人员基本信息、居住地信息、假释人员信息、收监人员信息、社区矫正定位信息、社区矫正转安置帮教信息这七大类信息。这些信息的具体内容如下：

（1）调查评估信息，即对被告人或罪犯信息的调查评估，一般由司法行政机关对被告人或罪犯所在乡镇、街道等进行调查，主要包括如下信息：被告人（罪犯）基本信息（姓名、身份证号、性别、出生年月、居住地地址、工作单位）、罪名、原判刑期、

原判刑期开始日期、原判刑期结束日期、原判刑罚、附加刑、判决机关、判决日期、委托单位、委托调查书等。

(2) 社区矫正人员基本信息，即存储在各省(区、市)相关信息系统中的社区矫正人员的基本信息。一般包括如下信息：社区矫正人员编号、管理对象类别、姓名、曾用名、性别、民族、证件类型、证件号码、出生日期、有无护照、护照号码、有无回乡证、回乡证号码、有无台胞证、台胞证号码、文化程度、健康状况、是否有传染病史、心理是否健康等。

(3) 居住地信息，即存储在各省(区、市)相关信息系统中的社区矫正人员的居住地信息。一般包括如下信息：社区矫正人员编号、申请时间、迁入地所在省(区、市)、迁入地所在地(市、州)、迁入地所在县(市、区)、迁入地(乡镇、街道)、迁入地明细、变更理由、司法所审核人、司法所审核时间、司法所审核意见、县(市、区)司法局审批人、县(市、区)司法局审批时间、县(市、区)司法局审批意见等。

(4) 假释人员信息，即存储在监狱信息管理系统中的假释人员信息，一般包括如下信息：社区矫正人员编号、姓名、身份证号、性别、出生年月、民族、原政治面貌、籍贯、家庭住址、学历、学位、从业状况、职称、特长、是否三假人员、是否三无人员、是否再犯罪、判决机关、判决书号、判决日期、罪名、刑种、原判刑期、刑期起始日期、刑期结束日期、附加刑、刑期变动、奖惩情况、主要犯罪事实、是否累犯等。

(5) 收监人员信息，即存储在监狱信息管理系统中。收监人员是指将被收监执行的社区矫正人员。收监人员信息一般包括社区矫正人员编号、姓名、申请时间、提请理由、提请依据、司法所申请人、司法所审核人、司法所审核时间、司法所审核意见、县(市、区)司法局审核人、县(市、区)司法局审核时间、县(市、区)司法局审核意见、处理意见等。

(6) 社区矫正定位信息，即存储在社区矫正人员定位系统中，包括社区矫正人员的地理位置信息、越界记录信息、停(关)机记录信息，具体包括社区矫正人员编号、矫正单位、矫正状态、经度、纬度、定位时间、是否越界等。

(7) 社区矫正转安置帮教信息，即社区矫正人员在矫正期满后，转入安置帮教(在各级政府领导下，由相关机关或社会力量对矫正人员的帮助、管理、教育的活动)系统中，该信息存储在安置帮教系统内，一般包括社区矫正人员编号、姓名、曾用名、性别、民族、身份证号、出生日期、文化程度、健康状况、原政治面貌、婚姻状况、原工作单位、联系电话、个人联系电话、户籍所在省(区、市)、户籍所在地(市、州)、户籍所在县(市、区)、户籍所在地(乡镇、街道)、户籍所在地明细、固定居住地所在省(区、市)等。

除了以上七类需集成的数据外，系统还需与人民法院、人民检察院、公安机关以及其他政府部门进行数据集成与交换。

4.2 信息系统数据集成的基本模式

经过多年的发展，在企业数据集成领域，已经积累了很多经验，有很多成熟的框架可以利用。目前采用的数据集成模式主要有联邦式、基于中间件模型和数据仓库等。

1. 基于联邦数据库系统的数据集成模式

联邦数据库系统(Federated Database System，FDBS)由一些彼此协作而又相互独立的半自治数据库系统构成。可以说，FDBS是异构单元数据库系统的集合，其中的这些数据库系统相互提供访问接口，以便分享数据。FDBS可以把不同分布的数据源整合为虚拟的数据源或者数据服务。整合后的数据可以被看作单一数据源，通过统一的访问方法(如JDBC、ODBC、NFS、SOAP)进行访问。

FDBS按集成度可分为两类：紧密耦合联邦数据库系统和松散耦合联邦数据库系统。紧密耦合式FDBS使用统一的全局模式，将各数据源的数据模式映射到全局数据模式上，从而解决了数据源间的异构性。松散耦合FDBS不提供统一的接口，可以通过统一的语言访问数据源，将很多异构性问题交给用户自己去处理，但必须解决所有数据源语义上的问题。

基于联邦数据库系统的数据集成模式可实现面向多个异构数据库系统的集成，通过各数据源之间的数据交换格式进行一一映射。其优点是容易实现、灵活性好；并可集成来自非关系型数据源(如电子邮件和文本文件)的数据。此外，数据的联邦式视图构建起来比数据仓库更快，更易于修改。但其缺点在于不适合大批量数据的集成。

2. 基于中间件的数据集成模式

基于中间件的数据集成模式也称为数据的逻辑集成。它通过在中间层提供一个统一的数据逻辑视图(虚拟的数据服务层)来隐藏底层的数据细节，使得用户可以把集成数据源看为一个统一的整体。用户可通过统一的全局数据模型来访问异构的数据库、遗留系统、Web资源等。中间件位于异构数据源系统(数据层)和应用程序(应用层)之间，向下协调各数据源系统，向上为访问集成数据的应用提供统一数据模式和数据访问的通用接口。各数据源的应用仍然完成它们的任务，中间件系统则主要集中为异构数据源提供一个高层次检索服务。

基于中间件的数据集成模式通过包装器(Wrapper)/协调器(Mediator)中介结构模式来实现。通过一个全局模式(Mediated Schema)集成各数据源。数据仍保存在相互独立的数据源中，用户查询在全局模式上进行。

包装器负责对各异构数据源的数据进行链接、转换，并将它们封装为统一的数

据模型。每种数据源对应一个包装器，对不同查询方式的新数据源，再构建一个与之匹配的包装器。

协调器提供统一的查询界面，对全局模式进行查询，根据数据源的元数据和映射规则将全局查询分解为对各个数据源的查询。然后把子查询发送到各个数据源的包装器中，并由协调器将结果进行整合集成返回给用户。在此过程中用户不需要知道数据源位置、模式等信息。

基于中间件的数据模式是目前比较流行的数据集成方法，其优势为：数据集成处理在中间件服务器进行，对数据的处理比较灵活，应用和底层的数据实现松耦合；当一个请求涉及多个底层数据源时，对底层的数据访问可以采用并发方式进行；借助中间件的灵活性，数据可以采用多种方式对外提供接口，从而大大方便各种应用的开发；所有数据都是实时从数据源取来，可保证数据的时效性。其缺点是：数据的处理在中间件层进行，由于从数据源到中间件层的数据传输开销，当数据量非常大时，效率比较低。其次，如何构造上述虚拟的逻辑视图，并使得不同数据源之间能映射到这个中间层，是应用开发者必须解决的难题。

ESB 提供的集成功能的中间件服务的组合可以很好地解决这个问题，只需要将数据集成封装为服务挂接在 ESB 上，而不用再考虑数据源之间的映射了。

3. 基于数据仓库的数据集成模式

数据仓库(Data Warehouse)是一个面向主题、集成、相对稳定、反映历史变化的数据集合，用于支持管理决策。它向用户提供用于决策支持的当前和历史数据，让用户更快、更方便地查询所需要的信息，进行决策支持。数据仓库中的数据是在对原有分散的数据库数据抽取、清理的基础上经过系统加工、汇总和整理得到的，必须消除源数据中的不一致性，以保证数据仓库内的信息是关于整个企业的一致的全局信息。

基于数据仓库的数据集成模式将各个异构数据源中的数据复制出来，通过分析、转换和封装等方法，将分散、不一致的数据转换为统一、同构的数据，存储在同一的数据仓库中，并保持数据源整体上的一致性。用户可以像查询本地数据库一样，访问想要的数据仓库数据。

数据仓库系统的优点有：

(1) 适合处理数据量庞大的集成应用；

(2) 能够提供很强的应用功能服务。

数据仓库系统的缺点有：

(1) 数据集成是面向主题的，集成目标明确，当主题改变时缺少适应性；

(2) 面向大型应用，集成模式不适合轻量级的简单应用；

(3) 基本上不具备实时处理能力，其数据抽取操作以定时方式从业务系统中

抽取。

上述基于联邦式、中间件和数据仓库构造的数据集成模式，可以从不同的着重点和应用层面解决数据共享问题。社区矫正信息系统的运行环境是基于SOA的云计算平台，具有数据量大、数据来源异构性强、数据分布式等特点。不仅要吸收上述三种数据集成技术的优点，还有探索适合基于SOA的云计算平台的数据集成新方法，例如基于SOA与ESB、基于REST的数据集成方法。

4.3 基于SOA与ESB的数据集成策略

4.3.1 SOA与ESB简介

SOA（Service-Oriented Architecture，面向服务的架构）是一种体系结构风格，又是一个组件模型，它将应用程序的不同功能单元-服务，通过服务间定义良好的接口和契约联系起来。其接口采用中立方式定义，独立于具体实现服务的硬件平台、操作系统和编程语言。构建在这种系统中的服务可使用统一和标准的方式进行通信和交互。

在SOA中所有业务流程都被定义为服务，服务通过基于类封装的服务接口委托给服务提供者，服务接口根据可扩展标识符、格式和协议单独描述。SOA提供了一种构建信息系统的标准和方法，并通过建立起综合、可重用的服务体系来减少IT业务冗余并加快项目开发的进程，使得开发部门效率更高、开发周期更短、项目分发更快。

应用SOA，可对不同时期、不同类型的异构系统以及跨企业边界的软件系统进行整合。基于SOA架构的应用集成可以减少不同类型的系统的依赖性，降低费用和操作的复杂性；提高已部署系统的灵活性，同时排除了抑制业务创新的障碍。

基于SOA的数据集成方法采用基于元数据和开放标准的共享服务方式，通过可重用设计和统一的方法实现数据的访问、集成和提交。其具备广泛的连接性，支持多种类型、结构和来源的企业数据集成，并且具备面向异构IT环境变化的应变能力。

SOA架构独立于技术实现，但通常通过Web服务实现SOA。Web服务是新一代的Web应用程序，它代表了组件技术和Web技术的结合，可远程而透明地调用和集成世界任何一个角落（运行在不同平台上）的一个服务。W3C定义：Web服务是一种通过URI标识的软件应用，其接口及绑定形式可通过XML标准定义、描述和检索，并能通过XML消息及互联网协议完成与其他应用的直接交互。

SOA中有三种角色：服务提供者、服务消费者和服务注册中心。其中服务提供者负责服务功能的具体实现，并通过注册服务操作将其所提供的服务发布到服

务注册中心，当接收到服务消费者的服务请求时，应执行所请求的服务。服务消费者是服务调用者，它首先到服务注册中心查找满足需求的服务，再根据服务信息进行服务绑定，以获得需要的功能。服务注册中心用来向服务提供者提供注册服务，以及对服务的分类和查找功能。

Web 服务可通过 Soap 技术或 REST 技术实现。基于 Soap 的 Web 服务涉及的关键技术包括：

(1) XML(Extensible Markup Language)——可扩展的标记语言，为 Web Service 提供了统一的数据格式，包括消息、服务描述以及工作流的描述。

(2) SOAP(Simple Object Access Protocol)——用于交换 XML 编码信息的轻量级协议。

(3) WSDL(Web Service Definition Language)——借助 XML 来描述一个网络服务或端点。用于定义 Web Service 以及调用方式。

(4) UDDI(Universal Description Discovery and Integration)——提供了在 Web 上描述并发现商业服务的框架，是面向 Web 服务的信息注册中心的实现标准和规范。

ESB(Enterprise Service Bus，企业服务总线)是 Web 服务、XML 等相关技术与传统的中间件技术相结合的一种产物。与计算机硬件总线类似，ESB 用于将企业内部或企业间实现不同功能的软件服务整合在一个统一的平台，从而向服务调用方提供统一的协议和接口，通过消息中介方式实现同构或异构系统的互联互通。ESB 还具有服务注册、服务生命周期、服务安全等管理以及通过服务组件或服务构件的方式实现基本服务、组合服务的组装，从而形成新的组合服务或流程服务，实现 SOA 的灵活、快速应变功能。

目前，ESB 已成为实现 SOA 应用架构的核心组件，为 SOA 架构提供结构更加稳定的、可维护性更强的、更统一的基础设施，实现分布式、松散耦合、基于事件驱动的 SOA 系统。ESB 对 SOA 架构中的服务进行统一管理，可向外部提供一组标准接口与调用方法，应用开发人员只需要将精力放在企业服务本身，而不需要关心服务如何被发现、调用，以及使用何种网络协议。ESB 的主要功能如下：

(1) 集中管理 SOA 中服务的元数据，对服务进行注册和路由寻址。

(2) 确保在 ESB 内的服务能进行可靠的消息传递。

(3) 统一管理在 ESB 中注册的服务，支持常用的传输协议，对外界提供统一的访问接口。服务调用者不需要知道所需服务具体部署的物理位置，从而简化了服务调用的过程。

(4) 支持多种服务的表现形式，如 Web 服务、消息、适配器等。

(5) 支持事务管理，保证对企业服务的原子性调用。

(6) 对服务调用提供日志和监控功能,如服务的调用日志、数据的传递方式等。

4.3.2 基于 SOA 与 ESB 的数据集成框架

基于 SOA 与 ESB 实现信息系统数据集成的基本思路是:将需要对外暴露的业务功能和数据存取功能封装成通用的 Web 服务,并发布到服务注册中心(UDDI),供其他 Web 服务调用者通过 ESB 提供的标准协议和数据规范进行查找和调用,在服务层实现集成数据。

ESB 作为中间件产品在各异构系统之间起着中介作用,成为 SOA 服务提供者和服务请求者之间的消息桥梁,对消息驱动和服务进行灵活的管理。各个应用系统通过 ESB 提供的接口把数据提交给 ESB 进行处理,之后,通过 ESB 把数据返回给请求服务的应用系统。

根据上述思路设计的基于 SOA 与 ESB 的社区矫正系统数据集成框架如图 4-1 所示。其中系统各层的具体功能如下:

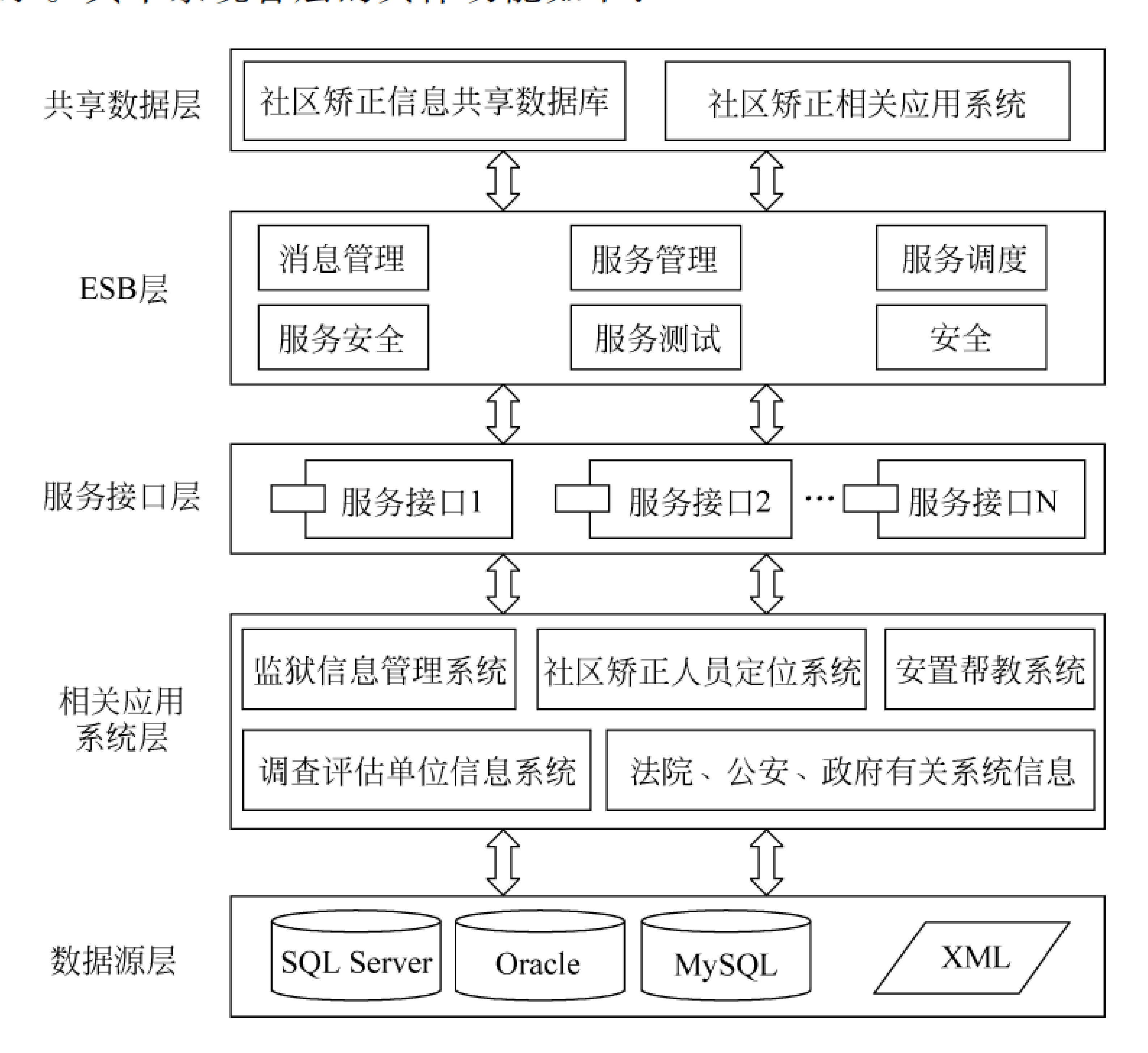


图 4-1 基于 ESB-SOA 的社区矫正系统数据集成框架

(1) 数据源层——存储了与社区矫正信息系统相关的原始数据。这些数据存储在与社区矫正相关单位、部门(司法行政、公安、人民法院、人民检察院等)的信息系统中,其存储格式可能是各种关系数据库(如 Oracle、Mysql、SQL Server)表、

Excel 表格、XML 数据文件等。

(2) 相关应用系统层——该层为与社区矫正相关的部门已建立的信息系统层，需要集成的社区矫正数据都来自这些系统。

(3) 服务接口层——包含与数据集成相关的所有 Web 服务接口，这些接口通过 VPN/政府专网等网络进行发布。位于“企业应用系统层”的信息系统可对这些服务接口进行调用，并将有关数据通过这些服务接口推送到本平台中，以待进一步处理。

(4) ESB 层——是应用系统集成框架的核心，负责对被集成的软件与数据进行处理，其主要功能包括：

- 消息管理——主要实现对消息的定义与处理。首先需要根据业务逻辑定义不同类型的消息，通过对消息类别的识别，实现对 Web 服务的调用。
- 服务管理——主要实现服务的声明、发布等操作。
- 服务调度——主要实现基于消息的路由，根据消息类型找到相应的 Web 服务处理程序。
- 服务安全——主要实现对 Web 服务调用者的身份验证、对所传输 XML 数据的加密，以及服务的安全调用。

(5) 共享数据层——集中存储、管理通过 ESB 集成得到的社区矫正信息共享数据库，供社区矫正相关信息系统，以及其他政府应用系统调用。

基于 ESB-SOA 的社区矫正系统数据集成框架中的运行过程为：首先把相关应用系统的相应软件模块以及遗留系统封装成 Web 服务，接着在 UDDI 注册中心注册、发布，并生成描述 Web 服务的 WSDL 文档。当服务请求者请求某项服务时，要先查找 UDDI 注册中心，寻找与该服务匹配的 WSDL 文件，如果找到，则根据相应 WSDL 信息与组件库中的服务进行绑定，调用服务，并返回结果。服务请求者发出的服务请求消息不是直接发送给服务提供者，而是将请求消息发送到 ESB，由 ESB 作为代理继续将请求消息转发给服务提供者。

4.4　基于 ROA 的数据集成策略

4.4.1　REST/ROA 基础

REST(Representational State Transfer，表示性状态转移)是一种充分利用 Web 特性的分布式软件架构风格，2000 年由美国的 Roy Thomas Fielding 在他的博士论文 *Architectural Styles and the Design of Network-Based Software Architecture* 中提出。Roy Thomas Fielding 是 Apache 基金会的第一任主席，是 HTTP 和 URI(Uniform Resource Identifier，统一资源标识符)协议的主要制定

者，参与过很多Web架构相关协议的设计，被称为Web的缔造者之一。REST是世界上最成功的分布式应用架构风格，HTTP /1.1就是为实现REST风格的架构而设计的。

REST从资源的角度来观察整个网络，整个Web被看作一组资源的集合。分布在各处的资源由URI确定，而客户端的应用通过URI来获取资源的表征。对资源进行的操作由HTTP协议动词的组合来实施。获得这些表征致使这些应用程序转变了其状态。随着不断获取资源的表征，客户端应用不断地在转变着其状态。

REST使用HTTP、URI、XML、HTML这些广泛流行的互联网协议和标准。其实现和操作较SOAP和XML-RPC简洁；性能、效率和易用性优于SOAP；可用缓存(Cache)机制提高响应的速度。随着Ajax、Ruby on Rails等Web开发技术的兴起，在Web开发技术社区掀起了一场重归Web架构设计本源的运动，REST架构风格得到了越来越多的关注，成为基于SOAP和Web服务描述语言(WSDL)的Web服务更为简单的替代方法。各种流行的Web开发框架，几乎都支持REST开发。主流Web 2.0服务提供者(Yahoo、Google、Facebook)采用了REST。

资源是REST中最关键的抽象概念，是能被远程访问的应用程序对象。一个资源就是一个标识单位，网络上任何可被访问或被远程操纵的事物(应用程序对象、数据库记录、算法、网页、图片等)都可被抽象成资源。资源可以是静态的，其状态永远不会改变；某些动态资源的状态可能随时间推移出现很大的变化。

资源的表示包括数据和描述数据的元数据，它代表资源当前状态的一些数据。同一个资源可存在不同的表示形式，例如订单的HTML表示可显示在网页浏览器中，XLS表示可以显示在Excel电子表格软件中，JSON表示可以为AJAX应用提供数据源，XML表示可以供应用程序解析处理。资源表示的多样性可以减弱对服务客户端的限制，增加了服务的松耦合性。

RESTful Web服务(也称为RESTful Web API)是使用HTTP并遵循REST原则的Web服务。它利用统一资源标识符(URI)来定位和识别资源，并通过HTTP协议中定义的标准方法(PUT、GET、POST、DELETE)对资源进行CRUD(创建、删除、查询、更新)操作。

通常，一个RESTful Web服务从以下三个方面来定义资源：

(1) 直观简短的资源地址(URI)，如http://example.com/resources/。

(2) 传输的资源：Web服务接收与返回的互联网媒体类型(如JSON、XML、YAML等)。

(3) 对资源的操作：Web服务针对资源的一系列请求方法(对资源的操作集合)，例如，POST、GET、PUT、DELETE等。

JSR 311或JAX-RS(用于RESTful Web Services的Java API)提供一组API，

以简化 RESTful Web 的开发。

ROA(Resource-Oriented Architecture,面向资源的架构)又称为 RESTful Architecture,是一种具体的 REST 架构,被称为 REST 式的 Web 服务架构。ROA 提供了一种把实际问题转换成 REST 式 Web 服务的方法,它将系统中所有事物都抽象成资源并赋予其唯一的资源标识符,通过 RESTful WebService 对外界提供各种资源服务,用一致的接口(如 URI(Uniform Resource Identifier,统一资源标识符))把资源暴露给外部世界,并提供对资源的操作服务。

ROA 作为 REST 式的 Web 服务构架,具有和 REST 同样的四大特征,即资源的可寻址性、无状态性、连通性和接口统一性。

可寻址性指资源通过 URI 暴露,而 URI 是可以寻址的。一个应用将其数据集中有价值的部分作为资源通过 URI 发布出来。

无状态性指每个 HTTP 请求都是无状态且完全孤立的,RESTful 服务端不需要在多次请求之间保留应用状态,服务端负责维护资源状态而不是应用状态,每一次操作不能影响到其他任何一个操作。这种无状态性使得每一个客户请求可由不同服务器来响应,当流量增加时,可通过添加新的服务器提高负载能力,如果某个服务器失效,可以直接将其从集群中移除,从而提高了系统的伸缩性。

连通性要求资源通过它们的表示彼此连接。通过资源的连通(链接),可以告知客户有哪些后续状态,例如网页中的“下一页”。

接口统一性是指对资源的各种操作必须通过统一接口实现,这不但能简化服务的开发和描述,而且在系统集成时便于服务的发现和自动匹配。HTTP/1.1 协议定义的统一接口包括四个通用操作接口,以及两个辅助性操作接口。

(1) 获取资源的一个表示:HTTP GET。

(2) 创建一个新资源:向一个新 URI 发送 HTTP PUT,或向一个已有的 URI 发送 HTTP POST。

(3) 修改已有资源:向已有 URI 发送 HTTP PUT。

(4) 删除已有资源:HTTP DELETE。

还有两个辅助性操作接口:

(1) 获取一个只包含元数据的表示:HTTP HEAD。

(2) 查看一个资源支持哪些 HTTP 方法:HTTP OPTIONS。

基于 SOAP 的 Web 服务的主要技术是 SOAP-WSDL-UDDI。其服务请求者将请求的服务封装成 SOAP 文件,发送到注册中心 UDDI,查找所需服务,获取服务描述文档 WSDL。之后请求者根据得到的 WSDL 完成与服务提供者之间的服务绑定,得到返回结果。这种方式效率低下,服务请求者每次调用复杂 Web 服务时,需要根据 UDDI 发给用户的 WSDL 文件反复进行 SOAP 信息交换,当大量的

服务请求同时进行时,可能造成网络堵塞或服务器瘫痪。

REST 建立在 Web 标准之上,直接使用 HTTP 协议,客户端和服务器都免除了解析和封装 SOAP 数据包的性能消耗,降低了传输负载。REST 通过 URI,直接定位到所需要的资源,不必通过繁冗的过程来获得资源,因此,在效率上优于 Web 服务。其次,REST 采用缓存机制来消除一些不必要的交互,也极大地提高了性能。可见,REST 有效降低了系统复杂度,其性能和效率优于基于 SOAP 的 Web 服务。

4.4.2 基于 REST/ROA 的数据集成框架

根据前述 REST/ROA 的基本原理,以及社区矫正信息系统数据集成的实际需求,我们设计出基于 REST/ROA 的社区矫正信息系统数据集成架构,如图 4-2 所示。

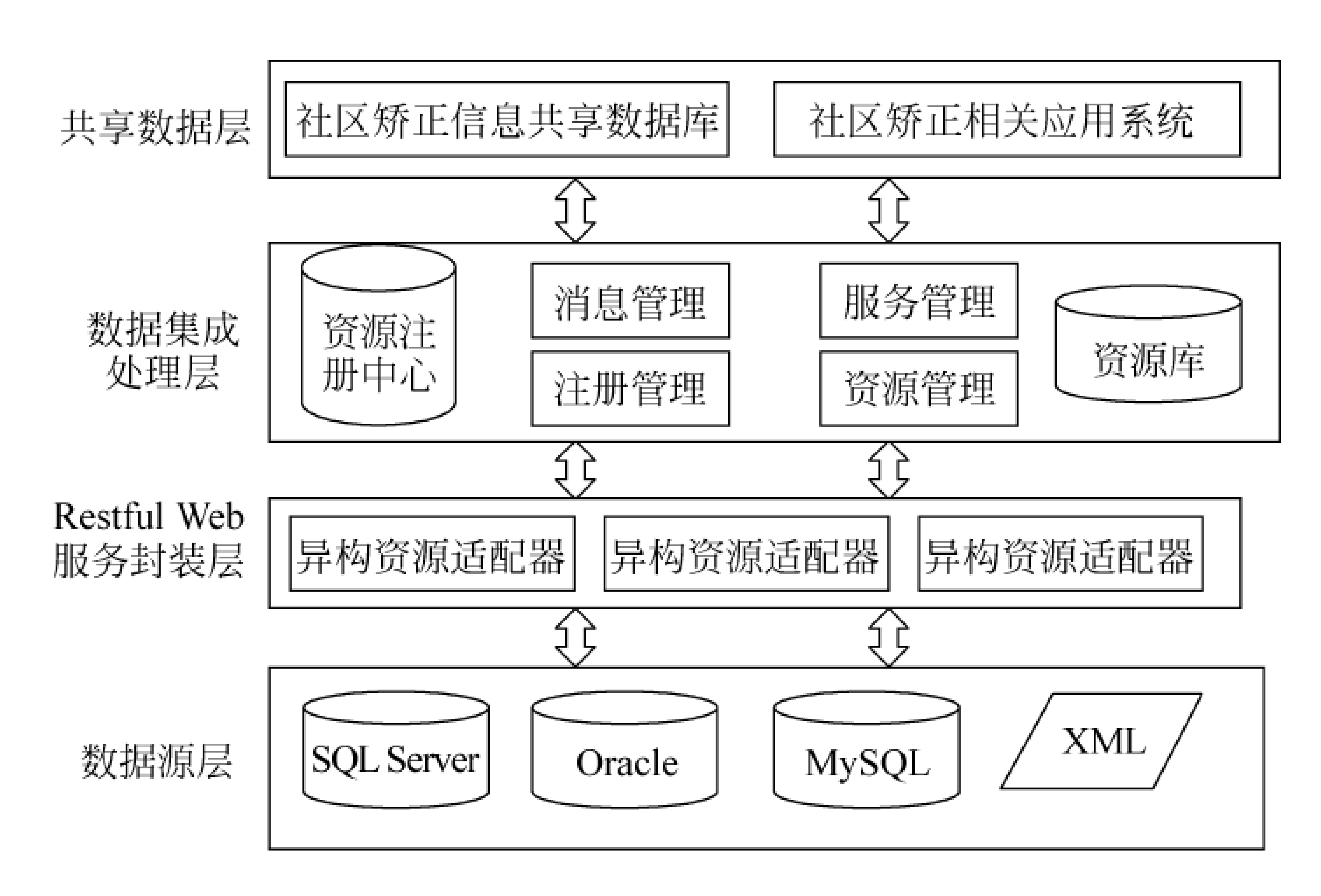


图 4-2 基于 REST/ROA 的社区矫正信息系统数据集成架构

下面介绍图 4-2 中核心模块的基本功能。

1. 数据源层

数据源层负责从异构系统获取、管理与社区矫正相关信息系统的原始数据。这些数据存储在与社区矫正相关单位、部门(司法行政、公安、人民法院、人民检察院等)的信息系统中。这些信息系统的运行环境可能是 Linux、Windows 或其他操作系统平台,其数据存储格式可能是各种异构关系数据库(如 Oracle、MySQL、SQL Server)表、Excel 表格、XML 数据文件等。

2. Restful Web 服务封装层

REST 采用面向资源的服务封装方法,向应用程序或者用户暴露的是资源而

不是具体的方法。本层通过多种资源适配器辅助产生数据源的各异构系统，把数据等业务资源封装为 Restful Web 服务，并将其注册到集成系统的资源注册中心，供数据集成层使用。

3. 数据集成处理层

本层是数据集成架构的核心，包括消息管理、服务管理、注册管理、资源管理、资源注册中心、资源库等。

- 消息管理模块：负责资源服务间消息转换和路由。包括资源服务的发布、订阅，消息检测与响应，服务调用历史记录等。
- 资源管理模块：分为资源模型管理与资源对象管理两大部分。

资源模型管理部分负责管理各异构系统的资源模型信息。资源模型的元信息和资源映射策略保存在 XML 文件中供资源对象部分使用。

资源对象管理部分管理各异构系统的资源对象信息。将服务模块的资源请求，通过资源策略管理器分派到对应资源处理引擎，将解析后的请求通过资源路由管理器从异构系统中获得资源表现，整合后返回给资源服务模块。

- 服务管理模块：管理资源服务的生命周期、发现和匹配、事务和安全。接收服务动态绑定请求，经过服务接口定义和配置后以 REST 服务文档格式发布，供外部应用调用。

4. 共享数据层

集中存储、管理通过 REST/ROA 集成得到的社区矫正信息共享数据库，供社区矫正相关信息系统，以及其他政府应用系统调用。

4.4.3 基于 REST/ROA 的数据集成子系统的设计步骤

根据 REST 规范，系统中所有可利用的数据及服务都是资源。在系统开发中，需要将资源按照合理设计的 URI 进行命名，利用 HTTP 提供的标准方法实现对资源的操作，实现数据到资源的转化，进而实现统一的对外访问接口。基于 REST/ROA 的社区矫正信息系统数据集成子系统设计步骤分为数据集规划、将数据集划分为资源、设计 URI 为资源命名、确定操作资源接口、设计客户端表示、设计服务器端表示、用超链接连通资源、考虑错误情况等。

1. 规划数据集

数据集指 ROA 将要对外暴露，或用户将要创建的数据集合。社区矫正相关系统以不同的数据结构、数据格式存储的大量数据就是数据集。根据需求，我们将社区矫正系统分成七个主要的数据集，分别是调查评估信息、社区矫正人员基本信息、居住地信息、假释人员信息、收监人员信息、社区矫正定位信息、社区矫正转安

置帮教信息数据集。

2. 将数据集划分为资源

确定系统的数据集后，需将数据集合理划分为资源，并作为 HTTP 资源进行发布。发布的 HTTP 资源主要有两种。一是专门预定义的一次性资源。例如信息系统首页，是系统其他资源入口的 URI。二是系统需要调用的数据资源。例如收监人员信息，它被暴露为资源，可以对其执行 GET、PUT 和 DELETE 等操作。

3. 使用 URI 为资源命名

资源的命名指为每一个资源赋予一个 URI（通用资源标识符）。URI 既是资源的名称，也是资源的地址，一个资源必须至少有一个 URI，而一个 URI 只能指示一个资源。URI 命名能清晰地描述资源。

4. 设计操作资源的接口

对社区矫正信息系统数据的操作，主要有新建、查询、更新、删除等操作。为了方便用户进行这些操作，需要通过网络对外暴露统一的操作接口。REST 直接使用 HTTP 方法，通过 HTTP GET、PUT、POST 或者 DELETE 方法来对请求资源进行相应的操作。当需要创建一个新资源时，可向该资源的 URI 发送 PUT 请求，或向已有资源发送 POST 请求；删除一个资源直接向该资源发送 DELETE 请求。

5. 设计客户端表示

在 ROA 中，客户端可通过浏览器实现。客户端表示指服务器发给客户端的资源表述，可以采用纯文本、XML、XHTML 等格式。目前使用较多的是 JSON 格式。

JSON（JavaScript Object Notation）是一种轻量级的数据交换格式，是基于 JavaScript（Standard ECMA-262 3rd Edition-December 1999）的一个子集。JSON 采用完全独立于语言的文本格式，能被 Java、Ruby 等程序读取并解析。JSON 易于人阅读和编写，同时也易于机器解析和生成。

6. 设计服务端表示

服务端表示指客户端请求服务器返回资源的数据表示格式，它必须能清楚地传达资源的当前状态和可能的下个状态的链接。客户端得到的表示包括当前资源状态和推进状态。客户端可以通过设置报头，例如设置 Accept，告知服务器返回自己需要的资源表示格式。服务器会通过阅读 Accept，按照客户端的要求格式返回给客户端。例如以 XML 文档格式返回给客户端，并转换为 Word 文档，以便工作人员下载存档。

7. 用超链接连通资源

遵循 REST 的连通性原则，用超链接和表单将新的资源与已有资源关联起来。

例如将系统页面中 form(表单)的 action(超链接)和资源关联起来。

8. 考虑出错情况

根据可能发生的出错情况,定义响应状态代码。主要考虑服务器和客户端发生的错误,这需要 HTTP 协议中的错误处理机制来完成。HTTP 的响应代码可以帮助服务器告知客户端哪些请求完成、哪些请求发生了错误。例如返回响应代码 200,表示请求顺利完成、返回。

4.5 基于 SOA 与 ESB 的数据集成方案

4.5.1 社区矫正信息系统数据集成思路

本系统需要进行集成的数据包括调查评估信息、社区矫正人员基本信息、居住地变更信息、假释人员信息、收监人员信息、社区矫正定位信息、社区矫正转安置帮教信息等。这些信息分别来源于不同的信息系统中,如表 4-1 所示。

表 4-1　需集成的信息及来源

需要集成的信息	信 息 来 源
调查评估信息	委托单位的信息管理系统
社区矫正人员基本信息	司法部社区矫正管理系统
居住地信息	司法部社区矫正管理系统
假释人员信息	监狱信息管理系统
收监人员信息	监狱信息管理系统
社区矫正定位信息	社区矫正人员定位系统
社区矫正转安置帮教信息	安置帮教系统
预留其他待集成信息	人民法院、人民检察院、公安机关相关系统

社区矫正数据集成系统的任务是在 VPN/公安专网/司法专网/政务专网的支持下,将分布在不同地点的相关信息系统数据集成到社区矫正系统中。图 4-3 描述了社区矫正数据集成系统的运行过程。

如图 4-3 所示,在基于 SOA 的数据集成平台中,通过 ESB 统一管理数据集成任务的消息与通信等操作。系统数据集成相关服务接口通过 VPN/专网发布,供社区矫正管理与行政机关、司法机关等信息系统进行调用,这些系统将所需的相关数据(以 XML 格式表示)消息推送至统一数据库。数据被推送消息后,由特定的消息捕获模块进行数据集成处理,将数据保存至社区矫正管理信息系统的统一数据库中,作为社区矫正管理信息系统的共享资源供有关软件模块调用。

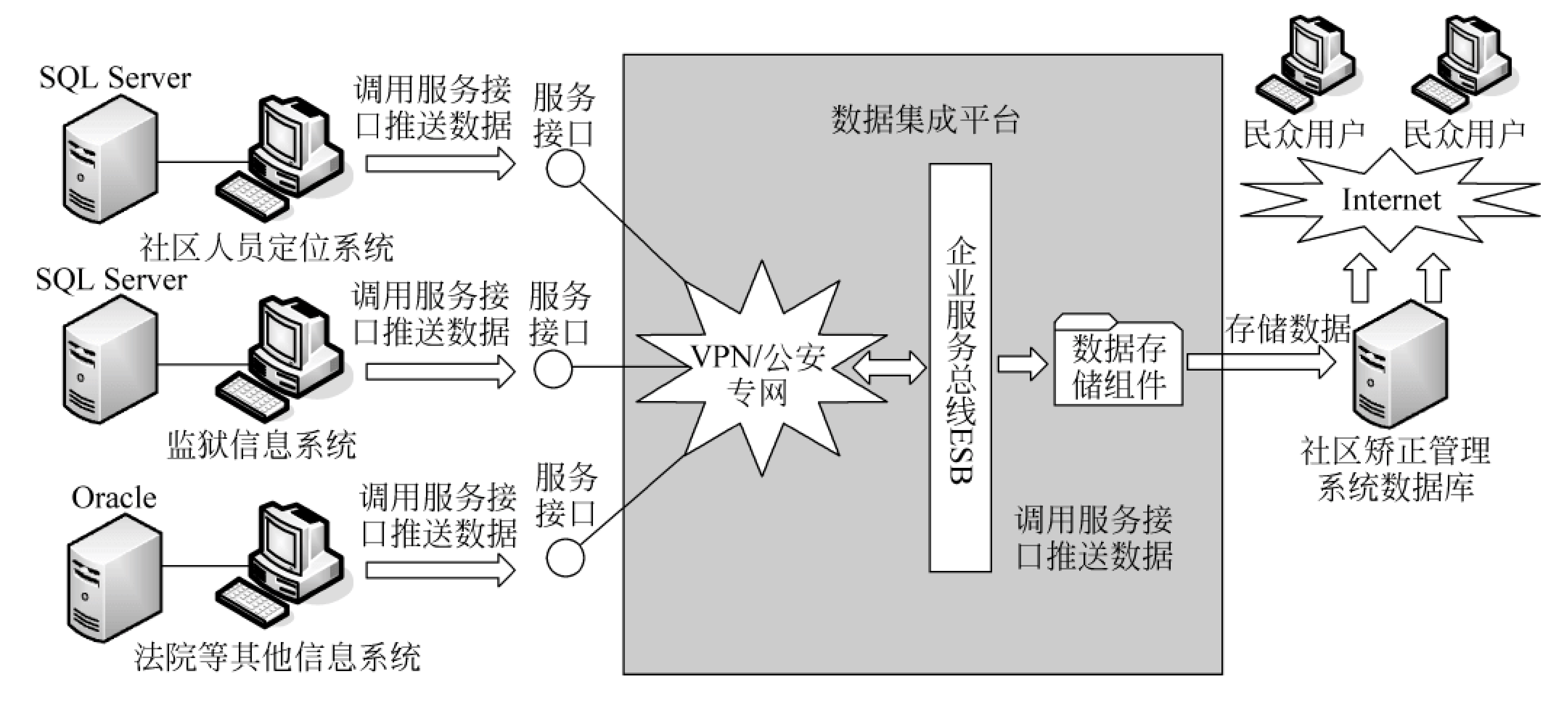


图 4-3 社区矫正数据集成系统方案示意图

4.5.2 社区矫正信息系统数据集成服务

根据需求分析社区矫正信息系统中主要包括如表 4-2 所示的七个数据集成服务，还需预留与人民法院、人民检察院、公安机关的数据交换服务接口。

表 4-2 实现的数据集成服务

序号	数据集成服务	服务的具体功能
1	调查评估信息的集成	通过委托单位的信息管理系统主动向社区矫正信息系统推送信息调查评估信息
2	社区矫正人员基本信息的集成	由基层单位(社区)采用主动推送的方式向社区矫正信息系统推送数据
3	居住地信息的集成	由基层单位(社区)采用主动推送的方式向社区矫正信息系统推送数据
4	假释人员信息的集成	由监狱信息管理系统主动推送数据至社区矫正信息系统，然后由社区矫正信息系统进行衔接转发或退回至监狱信息管理系统
5	收监人员信息的集成	社区矫正人员被收监执行时，社区矫正信息系统主动向监狱信息管理系统推送收监执行人员信息
6	社区矫正定位信息的集成	由社区矫正人员定位系统推送矫正对象人员的定位信息、越界记录等信息至社区矫正信息系统中
7	社区矫正转安置帮教信息	由社区矫正信息系统主动推送社区矫正人员信息至安置帮教系统内

4.5.3 数据集成软件模块结构

根据前述系统分析与系统架构设计的社区矫正信息系统数据集成子系统软件模块结构如图 4-4 所示。

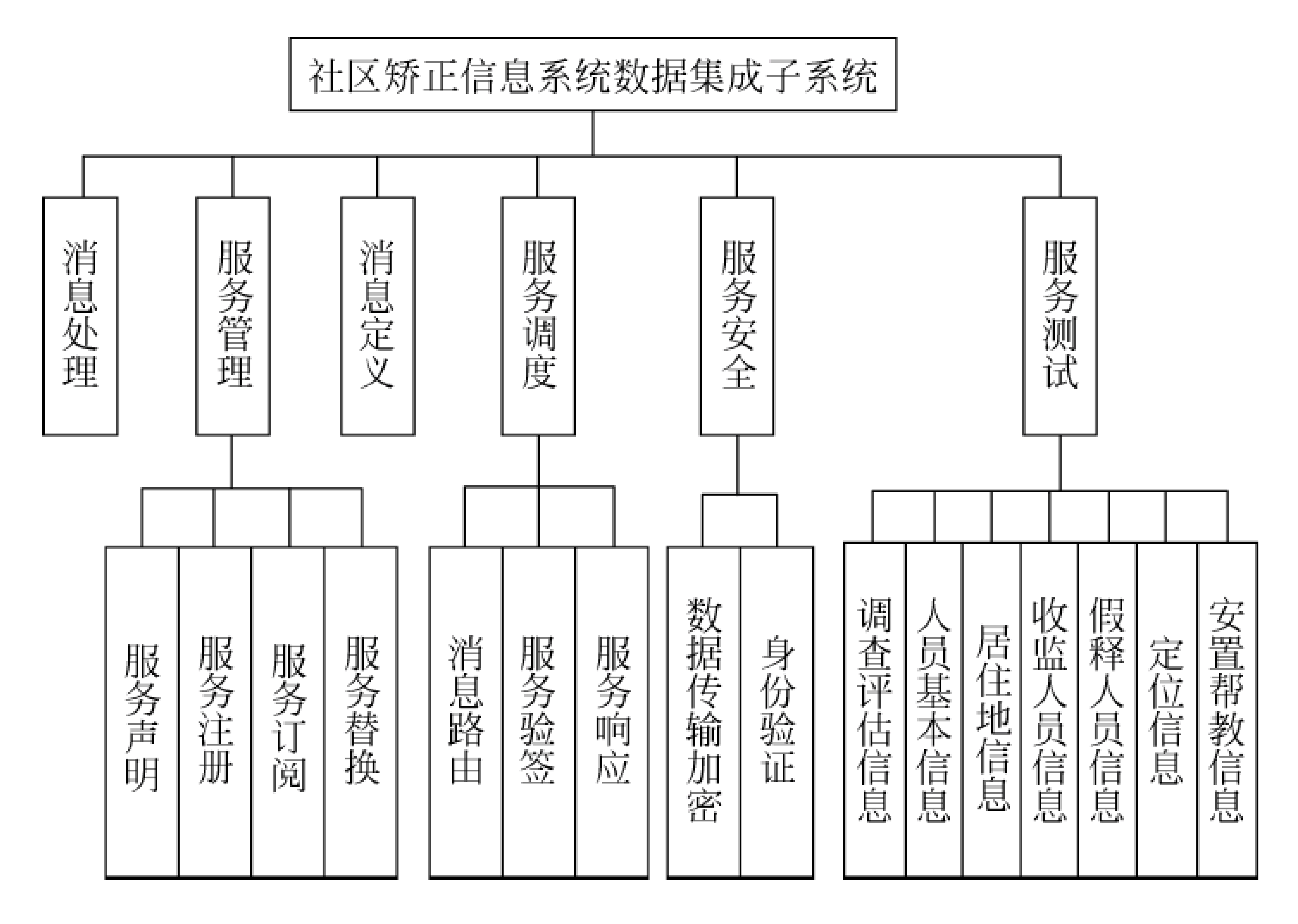


图 4-4　社区矫正信息系统数据集成子系统软件模块结构

图 4-4 中各主要模块的功能与实现步骤如下。

1. 消息处理模块

通过 ESB 框架 NServiceBus 进行消息的管理。消息处理的流程如下：

（1）从消息队列中查找是否有未经处理的消息。

（2）如果有，则启动一个分布式事务对其进行处理。

（3）从消息队列中提取一个消息。

（4）当消息被获得后，NServiceBus 尝试对该消息进行反序列化，如反序列化失败，则该消息被移动到一个专门存放错误消息的队列存放，然后该事务被提交。

（5）如果消息被成功反序列化，则 NServiceBus 调用相应的消息处理模块捕获该消息。如果在该步骤中出现了异常，则事务将会回滚，且接收的消息会重新返回到接收队列中。

2. 服务管理模块

服务管理模块主要对服务进行声明以及维护，为服务调度和业务运行提供服务支撑，实现流程如下：

首先需要对相关服务进行声明，然后由运维人员通过服务管理界面，将服务注册到服务平台上。经过对服务的测试验证，运维人员将可稳定运行的服务发布到

平台上。服务发布后，进入到服务目录中，供第三方业务系统进行订阅和发现。当服务组件过程或功能不满足业务要求时，可以将服务组件从服务平台注销掉。

以“社区矫正定位信息”的集成为例，该信息存放在Oracle数据库中，通过将Oracle数据库中得到的数据转换成为XML字符串，然后作为参数传递给本系统的Web服务，实现将Oracle数据库中的数据集成到本系统的SQL Server共享数据库中。

3. 消息定义模块

消息定义模块负责定义消息类型与消息主体，其实现流程如下：

(1) 定义消息的类型。

社区矫正信息系统包含七类数据集成服务，NServiceBus会根据消息的类型进行消息路由的选择，并最终找到相应的消息处理模块进行处理。社区矫正信息系统在消息的定义中包含一个枚举类型属性，指明当消息处理模块接收到该消息后，应进一步交给哪一个服务程序处理。

(2) 定义消息主体。

在NServiceBus框架中，通过实现IMessage接口来定义一个消息。

(3) 若消息具有时效性，则需要为消息类MyMessage添加时效属性。如果系统中对某一消息的处理程序不能在指定的时间内接收到该消息，则该消息将会因为超时而被丢弃。

(4) 对消息进行路由配置。

在NServiceBus框架中，消息的路由取决于配置文件中配置节MsmqTransportConfig的定义，需要在消息的发送方和接收方两个位置进行消息路由的配置。

4. 服务调度模块

服务调度模块主要根据业务请求，完成身份认证和服务解析，并根据消息类型进行消息路由，完成服务调用和处理，并返回服务调用结果。其处理步骤如下：

(1) 服务端接收到Web服务的调用。

(2) 得到Web服务定义中的消息类型。

(3) 根据消息类型进行判断，得到消息对应的服务程序。

(4) 对XML字符串进行解析，得到Web服务调用方的身份验证信息和待集成的数据列表。

(5) 使用第(4)步中得到的身份验证信息，在第(3)步的处理程序中验证该用户是否存在，若存在，则转到第(6)步，否则跳转到第(9)步。

(6) 使用第(4)步中得到的身份验证信息，在第(3)步的处理程序中验证该用户是否拥有集成此类数据的权限(消息类型中指明了集成数据的类型)，若存在，则

转到第(7)步,否则跳转到第(9)步。

(7) 调用系统的业务逻辑层对象,对待集成数据进行保存。

(8) 在第三方系统的客户端显示数据集成成功信息,流程结束。

(9) 在第三方系统的客户端显示数据集成失败信息,流程结束。

5. 服务安全模块

服务安全模块主要提供传输通道安全和数据安全,并提供用户签入授权与身份验证,确保服务平台应用信息安全。本系统主要通过以下两种安全策略保证系统的安全:

1) 数据传输安全策略

通过 RSA 算法对 xml 元素进行加密,保证数据传输的安全性。其加密的工作流程如下:

(1) 定位到要加密的 XML 元素。

(2) 使用 KeyInfo 元素中的密钥信息以及 EncryptionMethod 元素中的加密算法对该元素的内容进行加密。

(3) 对 KeyInfo 元素中的密钥使用公开密钥加密,并将加密后的密钥以及加密后的数据存放在 EncryptedData 元素中,并替换掉要加密的元素。

2) 身份验证策略

通过对用户进行身份验证和服务签入授权的两级验证方式,保证服务的安全性调用。身份验证策略的工作过程为:在客户端应用程序调用 Web 服务接口时,请求的 XML 格式字符串中需要加入 USER 和 PASSWORD 两个元素,代表社区矫正数据集成系统为该客户端指定的身份验证信息。当该 XML 字符串传输到数据集成系统后,系统会提取身份验证信息,并到数据库中进行核对。身份验证信息确定无误后,还需要进一步判断该用户是否拥有调用该接口的权限。社区矫正信息系统使用了基于角色的权限管理方式,其中用户和角色之间是多对多的关系。系统包含 7 类角色,分别能够集成 7 类社区矫正数据。新建用户后为其分配服务调用权限时,只需要将该用户 ID 与相应的角色 ID 建立起联系即可。

6. 服务测试模块

服务被声明并通过 Web 服务器 IIS 发布后,其他第三方信息系统,如监狱信息系统、社区矫正定位信息系统等就能对该服务进行引用并调用,其调用流程如下:

(1) 在客户端程序中添加 Web 服务的引用。

(2) 若是 VS 开发环境,则会由 VS 自动生成相应的.NET 中的托管类,由其他模块进行调用。

(3) 由客户端填写或选择相应的社区矫正定位信息数据。

(4) 根据填写的数据构造成相应的 xml 字符串参数。

(5) 调用社区矫正定位信息的 Web 服务接口。

(6) 根据调用的结果显示相应的成功或失败信息。

4.5.4 数据集成总体工作流程

基于如图 4-4 所示的社区矫正信息系统数据集成架构进行数据集成的工作流程如下：

首先需要定义 NServiceBus 框架内的消息、相关 Web 服务，并编写各个服务的响应程序，然后由第三方系统的客户端发送 Web 服务的调用请求，将相关数据推送到社区矫正信息系统中，实现社区矫正数据集成，具体工作流程如图 4-5 所示。

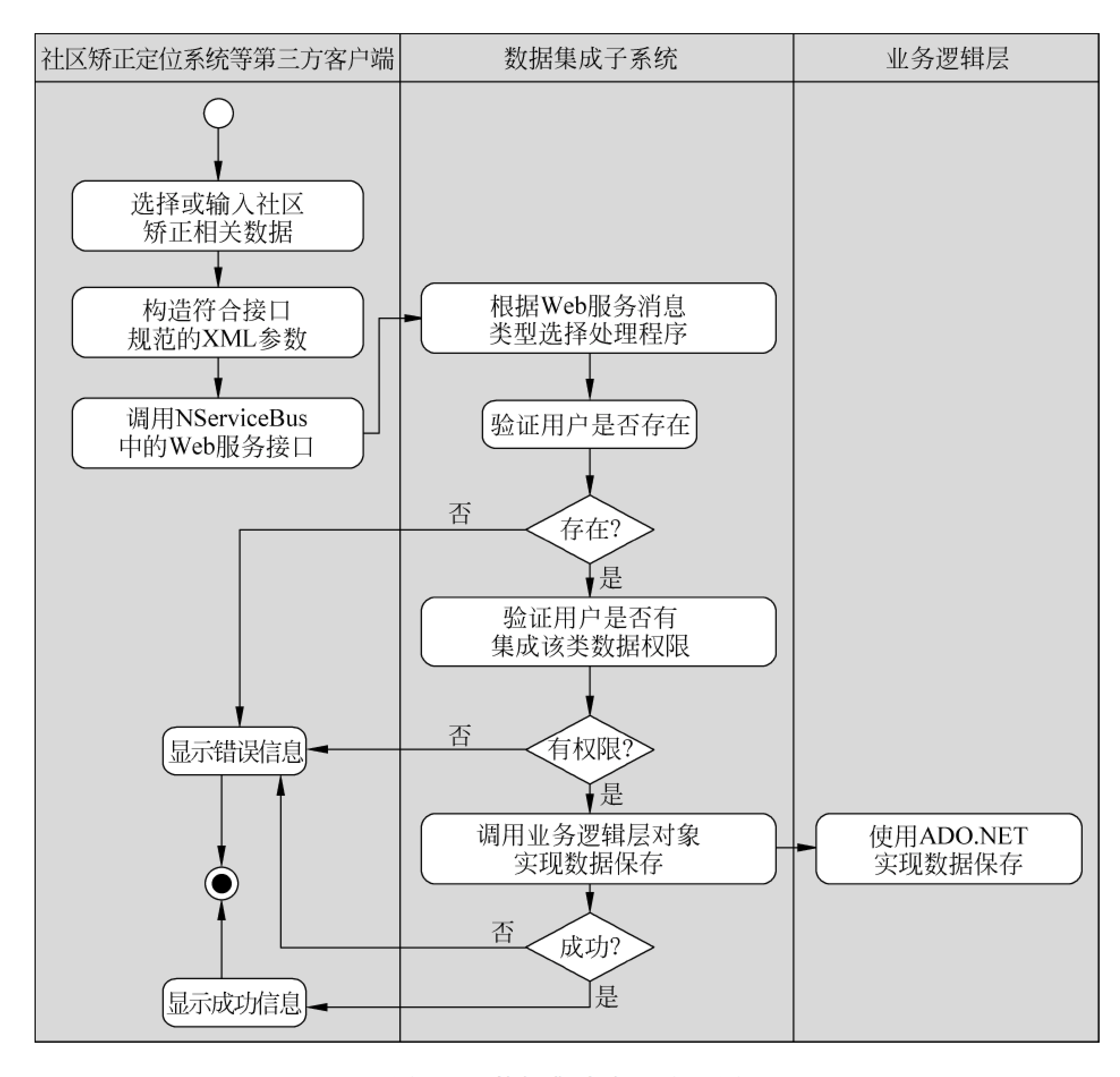


图 4-5 数据集成实现流程图

数据集成的操作由第三方系统（包括社区矫正定位系统、监狱信息系统等）发起，这些系统中存储着与社区矫正相关的数据。如图 4-5 所示的流程运行的具体

步骤如下：

（1）第三方系统经过查询或输入，得到待集成的社区矫正相关数据。

（2）将待集成的社区矫正数据，以及 Web 服务调用的身份验证信息构造成 xml 字符串，为 Web 服务的调用做准备。

（3）调用相应的 Web 服务接口。

（4）被调用的 Web 服务根据消息类型找到消息对应的处理程序。

（5）对 xml 字符串进行解析，得到 Web 服务调用方的身份验证信息和待集成的数据列表。

（6）使用第(5)步中得到的身份验证信息，验证该用户是否存在，若存在则转第(7)步，否则跳转到第(11)步。

（7）使用第(5)步中得到的身份验证信息，验证该用户是否拥有集成此类数据的权限(消息类型中指明了集成数据的类型)，若存在则转到第(8)步，否则跳转到第(11)步。

（8）调用系统的业务逻辑层对象，通过.NET 平台的 ADO.NET 实现对待集成数据的保存。

（9）通过 ADO.NET 组件的相关对象实现将数据保存到 SQL Server 2008 的相应数据表中，若成功则返回成功消息，否则跳转到第(11)步。

（10）在第三方系统的客户端显示数据集成成功信息，流程结束。

（11）在第三方系统的客户端显示数据集成失败信息，流程结束。

4.5.5　数据集成服务接口规范

由于 SOA 中服务的具体处理过程对服务调用者是透明的，一旦服务的具体处理方式发生了变化，只要服务接口不变，服务的调用者不需要做出任何改变即可应用新的处理方式。

本系统在设计几类数据集成服务的接口时，充分考虑到了这一点，力求接口的稳定。另一方面，由于 XML 已是网络中存储和传输数据的标准，具有良好的自描述的数据格式，因此在服务与服务调用者之间进行数据传递时，使用了 XML 数据格式。

基于 XML 的数据集成服务接口规范如下：

1. 调查评估信息接口规范

服务调用者在调用该服务时，构造如下的 XML 格式字符串，作为参数传递给服务接口：

```
<REQUEST>
<USER></USER>
```

```
<PASSWORD></PASSWORD>
<CONDITION>
<WTBH>委托编号</WTBH>
<NSYSQJZRYLX>拟适用社区矫正人员类型</NSYSQJZRYLX>
<BGRXM>被告人(罪犯)姓名</BGRXM>
<BGRSFZH>被告人(罪犯)身份证号</BGRSFZH>
<BGRXB>被告人(罪犯)性别</BGRXB>
<BGRCSRQ>被告人(罪犯)出生日期</BGRCSRQ>
<BGRJZDDZ>被告人(罪犯)居住地地址</BGRJZDDZ>
<BGRGZDW>被告人(罪犯)工作单位</BGRGZDW>
<ZM>罪名</ZM>
<YPXQ>原判刑期</YPXQ>
<YPXQKSRQ>原判刑期开始日期</YPXQKSRQ>
<YPXQJSRQ>原判刑期结束日期</YPXQJSRQ>
<YPXF>原判刑罚</YPXF>
<FJX>附加刑</FJX>
<PJJG>判决机关</PJJG>
<PJRQ>判决日期</PJRQ>
<WTDW>委托单位</WTDW>
<WTDCS>委托调查书</WTDCS>
</CONDITION>
</REQUEST>
```


数据集成服务在经过相应处理后，响应(返回结果格式)如下 XML 格式字符串给服务的调用者：


```
<RESULT>
<ERRORCODES>返回状态</ERRORCODES>
<DATA>
<RS>
<WTBH>委托编号</WTBH>
<BGRXM>被告人(罪犯)姓名</BGRXM>
<BDCRXM>被调查人姓名</BDCRXM>
<YBGRGX>与被告人(罪犯)关系</YBGRGX>
<DCSX>调查事项</DCSX>
<DCSJ>调查时间</ DCSJ>
<DCDD>调查地点</DCDD>
<NSYJZLB>拟适用矫正类别</NSYJZLB>
<DCDWSFS>调查单位(司法所)</DCDWSFS>
<DCDWXQJ>调查单位(县区级)</DCDWXQJ>
<DCR>调查人</DCR>
<DCYJSHR>调查意见审核人</DCYJSHR>
<DCPGYJ>调查评估意见书</DCPGYJ>
<CYQK>采用情况</CYQK>
</RS>
```

```
</DATA>
</RESULT>
```

其中返回状态为 0，表示操作成功；若为 1，则表示操作失败。

2. 社区矫正人员基本信息接口规范

由于本接口规范涉及 XML 元素过多，下面仅列出部分重要元素：

```
<REQUEST>
<USER></USER>
<PASSWORD></PASSWORD>
<CONDITION>
<SQJZRYBH>社区矫正人员编号</SQJZRYBH>
<GLDXLB>管理对象类别</GLDXLB>
<XM>姓名</XM>
<… …>… …</… …>
</CONDITION>
</REQUEST>
```

数据集成服务在经过相应处理后，响应（返回结果格式）如下 XML 格式字符串给服务的调用者：

```
<RESULT>
<ERRORCODES>返回状态</ERRORCODES>
</RESULT>
```

其中返回状态为 0，表示操作成功；若为 1，则表示操作失败。

3. 居住地变更信息接口规范

服务调用者在调用该服务时，构造如下的 XML 格式字符串，作为参数传递给服务接口：

```
<REQUEST>
<USER></USER>
<PASSWORD></PASSWORD>
<CONDITION>
<SQJZRYBH>社区矫正人员编号</SQJZRYBH>
<SQSJ>申请时间</SQSJ>
<QRDSZS>迁入地所在省(区、市)</QRDSZS>
<QRDSZD>迁入地所在地(市、州)</QRDSZD>
<QRDSZX>迁入地所在县(市、区)</QRDSZX>
<QRDXZ>迁入地(乡镇、街道)</QRDXZ>
<QRDMX>迁入地明细</QRDMX>
<BGLY>变更理由</BGLY>
<SFSSHR>司法所审核人</SFSSHR>
<SFSSHSJ>司法所审核时间</SFSSHSJ>
```

```
<SFSSHYJ>司法所审核意见</SFSSHYJ>
<XSFJSPR>县(市、区)司法局审批人</XSFJSPR>
<XSFJSPSJ>县(市、区)司法局审批时间</XSFJSPSJ>
<XSFJSPYJ>县(市、区)司法局审批意见</XSFJSPYJ>
</CONDITION>
</REQUEST>
```

数据集成服务在经过相应处理后，响应(返回结果格式)如下 XML 格式字符串给服务的调用者：

```
<RESULT>
<ERRORCODES>返回状态</ERRORCODES>
<DATA>
<RS>
<SQJZRYBH>社区矫正人员编号</SQJZRYBH>
<XM>姓名</XM>
<QRDXSFYJ>迁入地县(市、区)司法局意见</QRDXSFYJ>
</RS>
</DATA>
</RESULT>
```

其中返回状态为 0，表示操作成功；若为 1，则表示操作失败。

4. 假释人员信息接口规范

服务调用者在调用该服务时，构造如下的 XML 格式字符串，作为参数传递给服务接口：

```
<REQUEST>
<USER></USER>
<PASSWORD></PASSWORD>
<CONDITION>
<SQJZRYBH>社区矫正人员编号</SQJZRYBH>
<XM>姓名</XM>
<SFZH>身份证号</SFZH>
<XB>性别</XB>
<CSNY>出生年月</CSNY>
<MZ>民族</MZ>
<YZZMM>原政治面貌</YZZMM>
<JG>籍贯</JG>
<JTZS>家庭住所</JTZS>
<XL>学历</XL>
<XW>学位</XW>
<ZYZK>从业状况</ZYZK>
<ZC>职称</ZC>
<TC>特长</TC>
```

```
<SFSJRY>是否三假人员</SFSJRY>
<SFSWRY>是否三无人员</SFSWRY>
<SFZFZ>是否再犯罪</SFZFZ>
<PJJG>判决机关</PJJG>
<PJSH>判决书号</PJSH>
<PJRQ>判决日期</PJRQ>
<ZM>罪名</ZM>
<XZ>刑种</XZ>
<YPXQ>原判刑期</YPXQ>
<XQQSRQ>刑期起始日期</XQQSRQ>
<XQJSRQ>刑期结束日期</XQJSRQ>
<FJX>附加刑</FJX>
<XQBD>刑期变动</XQBD>
<JCQK>奖惩情况</JCQK>
<ZYFZSS>主要犯罪事实</ZYFZSS>
<SFLF>是否累犯</SFLF>
</CONDITION>
</REQUEST>
```

数据集成服务在经过相应处理后，响应（返回结果格式）如下 XML 格式字符串给服务的调用者：

```
<RESULT>
<ERRORCODES>返回状态</ERRORCODES>
</RESULT>
```

其中返回状态为 0，表示操作成功；若为 1，则表示操作失败。

5. 收监人员信息接口规范

服务调用者在调用该服务时，构造如下的 XML 格式字符串，作为参数传递给服务接口：

```
<REQUEST>
<USER></USER>
<PASSWORD></PASSWORD>
<CONDITION>
<SQJZRYBH>社区矫正人员编号</SQJZRYBH>
<XM>姓名</XM>
<SQSJ>申请时间</SQSJ>
<TQLY>提请理由</TQLY>
<TQYJ>提请依据</TQYJ>
<SFSSQR>司法所申请人</SFSSQR>
<SFSSHR>司法所审核人</SFSSHR>
<SFSSHSJ>司法所审核时间</SFSSHSJ>
<SFSSHYJ>司法所审核意见</SFSSHYJ>
```

```
<XSFJSHR>县(市、区)司法局审核人</XSFJSHR>
<XSFJSHSJ>县(市、区)司法局审核时间</XSFJSHSJ>
<XSFJSHYJ>县(市、区)司法局审核意见</XSFJSHYJ>
<CLYJ>处理意见</CLYJ>
</CONDITION>
</REQUEST>
```


数据集成服务在经过相应处理后,响应(返回结果格式)如下 XML 格式字符串给服务的调用者:


```
<RESULT>
<ERRORCODES>返回状态</ERRORCODES>
</RESULT>
```


其中返回状态为 0,表示操作成功;若为 1,则表示操作失败。

6. 社区矫正定位信息接口规范

服务调用者在调用该服务时,构造如下的 XML 格式字符串,作为参数传递给服务接口:


```
<REQUEST>
<USER></USER>
<PASSWORD></PASSWORD>
<CONDITION>
<SQJZRYBH>社区矫正人员编号</SQJZRYBH>
<JZDW>矫正单位</JZDW>
<JZZT>矫正状态</JZZT>
<JD>经度</JD>
<WD>纬度</WD>
<DWSJ>定位时间</DWSJ>
<SFYJ>是否越界</SFYJ>
</CONDITION>
</REQUEST>
```


数据集成服务在经过相应处理后,响应(返回结果格式)如下 XML 格式字符串给服务的调用者:


```
<RESULT>
<ERRORCODES>返回状态</ERRORCODES>
</RESULT>
```


其中返回状态为 0,表示操作成功;若为 1,则表示操作失败。

7. 社区矫正转安置帮教信息接口规范

服务调用者在调用该服务时,构造如下的 XML 格式字符串,作为参数传递给

服务接口：


```
<REQUEST>
<USER></USER>
<PASSWORD></PASSWORD>
<CONDITION>
<SQJZRYBH>社区矫正人员编号</SQJZRYBH>
<XM>姓名</XM>
<CYM>曾用名</CYM>
<XB>性别</XB>
<MZ>民族</MZ>
<SFZH>身份证号</SFZH>
<CSRQ>出生日期</CSRQ>
<WHCD>文化程度</WHCD>
<JKZK>健康状况</JKZK>
<YZZMM>原政治面貌</YZZMM>
<HYZK>婚姻状况</HYZK>
<YGZDW>原工作单位</YGZDW>
<LXDH>联系电话</LXDH>
<GRLXDH>个人联系电话</GRLXDH>
<ZP>照片</ZP>
<HJSZS>户籍所在省(区、市)</HJSZS>
<HJSZDS>户籍所在地(市、州)</HJSZDS>
<HJSZXQ>户籍所在县(市、区)</HJSZXQ>
<HJSZDXZ>户籍所在地(乡镇、街道)</HJSZDXZ>
<HJSZDMX>户籍所在地明细</HJSZDMX>
<GDJZDSZS>固定居住地所在省(区、市)</GDJZDSZS>
<GDJZDSZDS>固定居住地所在地(市、州)</GDJZDSZDS>
<GDJZDSZXQ>固定居住地所在县(市、区)</GDJZDSZXQ>
<GDJZDXZ>固定居住地(乡镇、街道)</GDJZDXZ>
<GDJZDMX>固定居住地明细</GDJZDMX>
<ZM>罪名</ZM>
<YPXF>原判刑罚</YPXF>
<FJX>附加刑</FJX>
<YPXQ>原判刑期</YPXQ>
<SQJZJDJG>社区矫正决定机关</SQJZJDJG>
<YJYCS>原羁押场所</YJYCS>
<JZLNR>禁止令内容</JZLNR>
<JZLKSRQ>禁止令开始日期</JZLKSRQ>
<JZLJSRQ>禁止令结束日期</JZLJSRQ>
<JZLB>矫正类别</JZLB>
<SQJZQX>社区矫正期限</SQJZQX>
<SQJZKSRQ>社区矫正开始日期</SQJZKSRQ>
<SQJZJSRQ>社区矫正结束日期</SQJZJSRQ>
<FLWSSDSJJZL>法律文书收到时间及种类</FLWSSDSJJZL>
```

```
<JSFSJBDSJ>接受方式及报到时间</JSFSJBDSJ>
<ZYFZSS>主要犯罪事实</ZYFZSS>
<BCFZQDWFJL>本次犯罪前的违法犯罪记录</BCFZQDWFJL>
</CONDITION>
</REQUEST>
```

数据集成服务在经过相应处理后，响应（返回结果格式）如下 XML 格式字符串给服务的调用者：

```
<RESULT>
<ERRORCODES>返回状态</ERRORCODES>
</RESULT>
```

其中返回状态为 0，表示操作成功；若为 1，则表示操作失败。



社区矫正系统信息安全技术

社区矫正工作包括从衔接到解除的全过程的管理和监督、社区矫正人员的定位、社区矫正人员信息的综合查询、统计分析等，涉及人民法院、人民检察院、公安机关、监狱等方面的用户对拟适用社区矫正的被告人、罪犯进行信息管理和监管。由于系统涉及各级人民法院、人民检察院、公安机关等用户，并且需要与法院（预留）、公安机关（预留）、监狱部门的业务系统对接，实现社区矫正人员信息的自动交换，因此，为了保证社区矫正管理系统的信息的准确性、完整性、可靠性、可信性、保密性，必须建立完善的安全机制。

5.1　信息安全技术基础

信息系统的安全需求包括系统信息的准确性、完整性、机密性，也包括用户身份的可鉴别性及资源访问的可控制性，这些安全需求的实现需要依赖于基础的信息安全技术，本节重点阐述在构建信息系统安全框架时需要的几种信息安全基础技术。

5.1.1　信息的机密性保障技术

在信息系统中传输的敏感信息需要提供机密性安全措施，这主要通过信息保密技术实现。

1. 信息保密技术概述

在开放的网络环境中，即使截获的是密文信息，密码分析者也可以通过分析，

从截获的密文推断出原来的明文或密钥。其中主要的窃取方式包括被动攻击(Passive attack)和主动攻击(Active attack)。被动攻击是指一个保密系统采取对密文截获并进行分析的攻击方式；主动攻击是指恶意入侵者(Tamper)、攻击者(Attacker)或黑客(Hacker)主动对系统窜扰，利用删除信息、增添信息、信息重放、信息伪造等手段向系统发起攻击，以达到自己的目的。

由信息安全的攻防两个方面共同构筑了信息保密系统的模型，如图 5-1 所示。

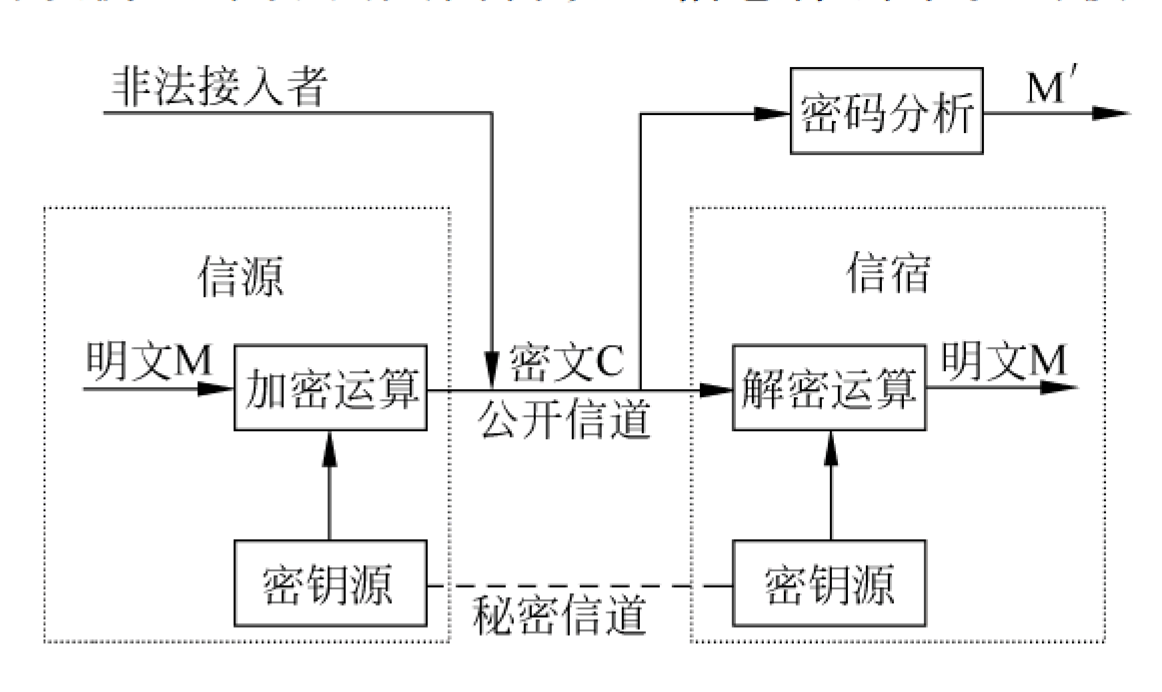


图 5-1　信息保密系统模型

密码系统从原理上可以分为两大类，即单钥体制(One-key system)和双钥体制(Two key system)。单钥体制是指加密密钥和解密密钥相同，而双钥体制是指加密密钥和解密密钥不同。

单钥体制也称为对称密码体制，这种密码体制在应用中的一个主要问题是：如何将加密密钥通过秘密信道分发给消息的接收者，即密钥的生成和管理问题。对称密码体制根据加密方式的不同又分为流密码(Stream cipher)和分组密码(Block cipher)。其中，流密码是对于明文按照字符逐位加密的方式完成加密；而分组密码则是先对明文消息分组，再逐组完成加密。

双钥体制也称为非对称密码体制或公钥体制(Public key system)，是由于 Diffie 和 Hellman 在 1976 年开创性的工作而诞生的。在该体制中，每个用户都有一对选定的密钥(公钥 PK 和私钥 SK)，公开的密钥 PK 可以像电话号码一样进行注册公布，而私钥 SK 则由用户秘密持有。非对称密码体制所基于的数学基础必须保证从公开密钥推出私钥是不可行的。非对称密码体制的主要特点是，可以实现多个用户加密的消息只能由一个用户解读(用于公共网络中实现保密通信)。相对于对称密码体制，非对称密码体制的一个优越性是无须事先分配密钥。

对称密码体制和非对称密码体制各有其优缺点和适应性。对称密码体制的加解密算法效率高，因此被用于大量消息的加密，而其应用的一个突出问题是密钥如何安全的分发，因为在开放的网络环境中要找到一个可用于密钥传输的安全信道是非常困难的；非对称密码体制的突出优点是没有密钥分发问题，用于解密的公

钥可以完全公开，其双钥特点还可以使其用于对消息的数字签名。

2. 对称密码体制

对称密码又分为流密码和分组密码。流密码的主要思想是以尽可能简单的方式来生成一个伪随机性尽可能好的周期序列。流密码体制以简洁、快速的生成算法，使其成为新一代移动通信的主流加密算法；分组密码是将明文序列划分成等长的分组，对每一组用同一加密算法和同一密钥进行加密。分组密码体制具有简洁、快速的特点，并且容易实现标准化，使其成为主流的软硬件加密标准。

分组密码是现代密码学的重要分支之一，其主要任务是提供数据保密性。分组密码算法采用一些固定的置换对明文数据分组进行加密变换，首先将明文分成固定比特长度的分组，加密运算的输入为明文组和密钥组，经过加密变换得到密文组；解密算法是将密文组和密钥组经过变换得到明文组，其中，解密运算是加密运算的逆运算。

分组密码实质是要设计一种算法，在密钥控制下，把 n 比特明文置换成唯一的 n 比特密文，并且这种加密变换必须是可逆的，即密文通过解密运算必须能够解密得到明文。

分组密码算法的设计思想是由 C. E. Shannon 提出的，主要通过扩散(diffusion)和混淆(confusion)两种手段来实现。扩散的目的是将明文组和密钥组的影响扩散到密文组中，常常通过“置换(Permutation)”的方法来实现扩散；混淆的目的在于使密钥和密文之间的关系变得复杂，以实现明文与密文之间以及密文与密钥之间具有极小的统计相关性，从而使统计分析类型的攻击难于奏效，常常使用“代换(Substitution)”的方式以达到混淆的目的。

分组密码有两个重要的参数：一个是密钥的比特位数，称为密钥长度；另一个是分组的比特位数，称作分组长度。用 $E_k(m)$表示密钥 k 对明文 m 加密，$D_k(c)$表示密钥 k 对密文 c 的解密，显然，必然有式(5.1)成立。

$$D_k(E_k(m)) = m \tag{5.1}$$

分组密码的设计要求主要有以下几点：一是分组长度足够长(一般为 65～128 比特)；二是密钥长度要足够长(65～128 比特)；三是算法足够复杂，包括加解密算法和子密钥生成算法；四是加密、解密算法易于软件和硬件的实现；五是便于分析，即算法简洁清晰但破译困难。

下面介绍两种最常用的分组加密算法。

1) DES 算法

DES 密码算法是 1977 年由美国国家标准局公布的第一个分组密码算法。1973 年，为了建立适用于计算机系统的商用密码，美国商业部所属国家标准局 ANBS 开始研究应用于国防部之外的其他部门计算机系统的数据加密标准，并于

1973 年 5 月及 1974 年 8 月两次向公众发出公告，征求加密算法。在征得的算法中，IBM 公司提出的 Lucifer 算法入选。DES 密码实际上是 Lucifer 密码的进一步改进。1975 年 3 月，ANBS 在联邦记录中公布了 DES 算法，1977 年 1 月正式向社会公布 DES 算法，作为非机密数据的正式数据加密标准（Data Encryption Standard，DES）。

DES 自 1977 年由美国国防部采用后，每隔五年由美国国家保密局（NSA）对 DES 做出一次评估，并重新批准它是否继续作为联邦加密标准。DES 在国际通信保密舞台上活跃了 25 年后，在 21 世纪初被新的数据加密标准 AES 取代。

DES 加密运算包括一个初始置换 IP，16 个轮运算以及一个末置换 IP^{-1}，最后得到 64 比特的输出。首先对 64 比特的输入进行初始置换 IP，得到 64 比特的输出；在 DES 加密的 16 个轮运算中，每一轮的输入是上一轮输出的 64 比特及 48 比特子密钥进行运算的结果，轮输出为 64 比特。其中每轮所使用的 48 比特子密钥是由密钥调度算法根据 56 比特的初始密钥一次产生；完成 16 轮运算之后，输出结果被分成两半对调，然后进行末置换 IP^{-1}。以上描述的 DES 算法的基本结构如图 5-2 所示。

从图 5-2 中可以看出，DES 算法的核心轮运算共 16 轮，每轮中包括一次 F 运算及一次异或运算。其中第 i 轮的 F 运算需要子密钥 K_i 参与，轮迭代中 F 运算的具体细节参见相关参考文献。

DES 算法 1977 年首次公之于世以来，学术界对 DES 密码进行了深入的研究，特别是对其安全性和破译方法深入的研究推动了密码学理论的发展。人们一直对 DES 的安全性持怀疑态度，对 DES 的批评主要集中在以下三个方面。

第一，密钥的长度。作为分组密码，DES 的密钥仅有 64 比特，其中某些位还要用于奇偶校验或其他通信开销，有效位只有 56 比特，这对于数据安全性来说显得不足。

第二，轮子密钥之间的相关性。各次迭代中使用的轮子密钥是由有效位为 56 比特的密钥递推产生的，在递推过程中，各轮子密钥之间会产生相关性，这在一定程度上降低了密码算法的安全性。

第三，S 盒的设计。DES 迭代中唯一的非线性运算部件 S 盒的设计原理尚未完全公开，其中可能留有隐患。更有人担心 DES 算法中有“陷门”，知道秘密的人可以很容易对密文进行解密。

2）高级加密标准 AES

从 1976 年 DES 算法公布以来，到 20 世纪末，该算法基本主宰了对称加密算法的研究和开发。随着密码分析技术、芯片处理能力以及计算技术的不断进步，专家们普遍认为，DES 算法及其变形的安全性难以满足新的应用需求，可能存在安全

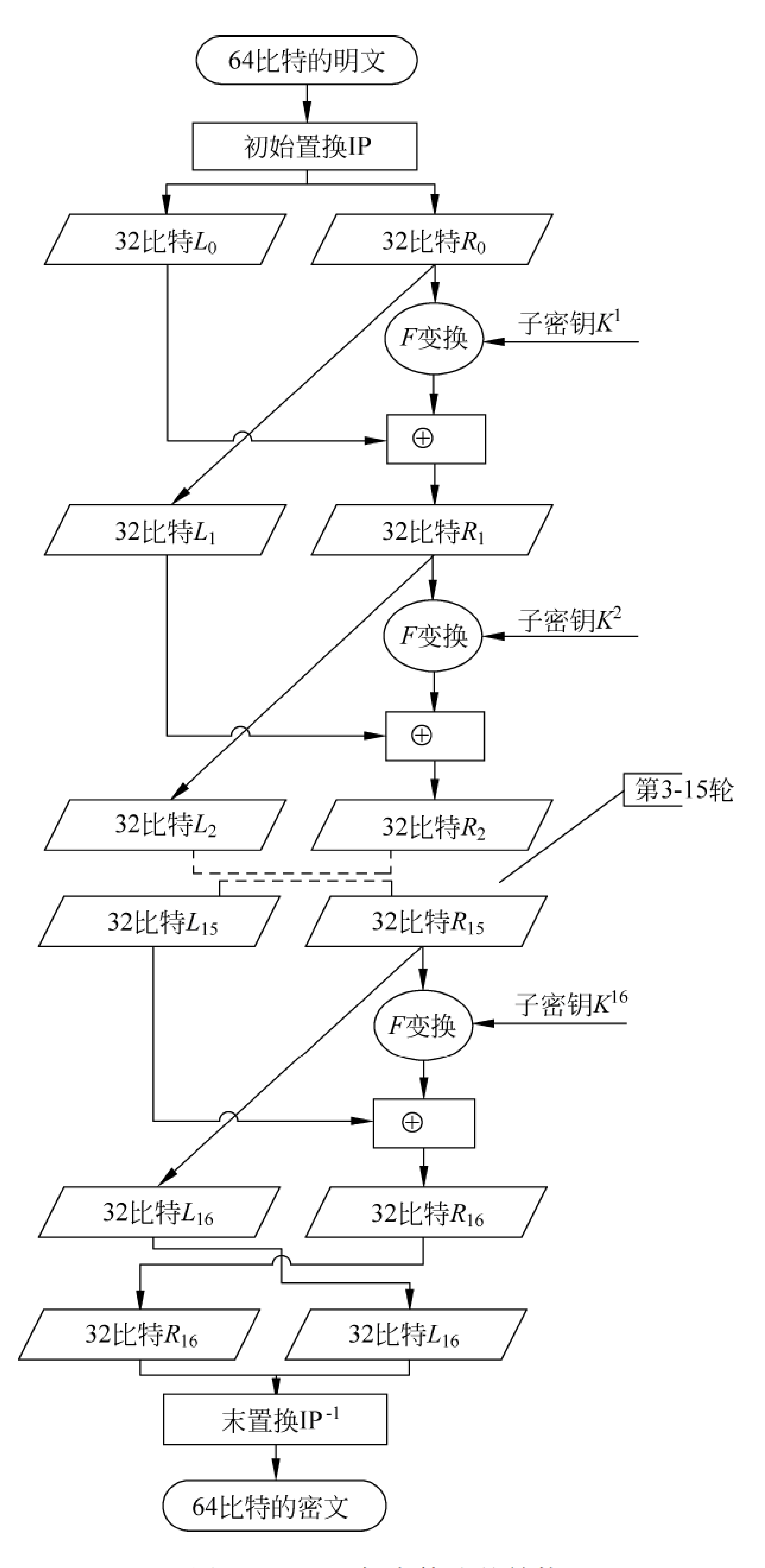


图 5-2　DES 加密算法的结构

方面的隐患，因此迫切需要一种安全性更高的算法作为新一代的分组加密标准。美国国家标准与技术研究所（NIST）提出新的对称加密标准 AES（Advanced Encryption Standard），并于 1997 年公开征集算法。Rijndael、Twofish、Mars、

RC6、Serpent 五个算法进入第二轮筛选，经过几年的专家评审，测试和反复论证，两名比利时研究者 Vincent Rijmen 和 Joan Daemen 设计的 Rijndael 算法因其在安全性、性能、实现特性等方面的优势，而在 2001 年被正式发布作为 AES 标准。

Rijndael 是分组长度和密钥长度均可变的分组密码算法，密钥长度和分组长度可以独立地指定为 128 比特、192 比特或 256 比特，分别记为 AES-128、AES-192、AES-256。

AES 加密算法示意图如图 5-3 所示。Rijndael 算法在整体结构上采用 Square 结构，这是一种类似幻方的多轮迭代结构。每一轮由三层组成。

(1) 非线性层：进行 S-盒变换 SubByte，起到混淆的作用。

(2) 线性混合层：进行行移位变换 ShiftRow 和列混合变换 MixColumn，以确保多轮之上的高度扩散。

(3) 密钥加层：进行密钥加变换 AddRoundKey，将轮密钥简单地异或到中间状态上。

Rijndael 算法以字节(8 比特)，字(32 比特)为处理单位，将明文分为 N_b 个字，密钥分为 N_k 个字，每个字为 5 个字节。算法共进行一个初始轮和 N_r-1 轮变换及末轮变换。

对称密码体制在实际应用中的一个问题是：如何实现密钥的分发，即由于加密和解密使用的是相同的密钥，在一个开放的网络环境中，如何安全地在信息的发送方和信息的接收方之间实现密钥共享的问题。

3. 非对称密码体制

非对称密码体制(Asymmetric cryptosystem)，也称为公钥密码体制(Public key cryptosystem)，是现代密码学的重要组成部分。非对称密码体制的思想在 1976 年由 Diffie 和 Hellman 在其“密码学新方向”一文中提出。Rivest、Shamir 和 Adleman 在 1978 年提出了首个非对称密码体制，即著名的 RSA 公钥密码体制。非对称密码体制的提出是现代密码学的具有里程碑意义的重要事件，它的出现标志着现代密码学的开始。

非对称密码体制与所有以前的密码方法有很大的不同。一是所基于的基本思想不同。以前的密码算法都基于代换与置换的基本操作，而非对称密码体制是使用数学函数进行变换的。二是密钥的使用方式的不同。传统密码算法仅使用一个密钥，其加密密钥和解密密钥是完全一致的；而公钥密码体制使用两个密钥(加密密钥和解密密钥)，其中一个密钥用于加密，该密钥可以对公众公开，也称为公开密钥；另外一个密钥用于解密，该密钥必须保密，也称为私有密钥。正是这一特点，使得非对称密码体制很好地解决了开放网络环境中的密钥分发问题。

非对称密码算法的密钥具有如下特点：

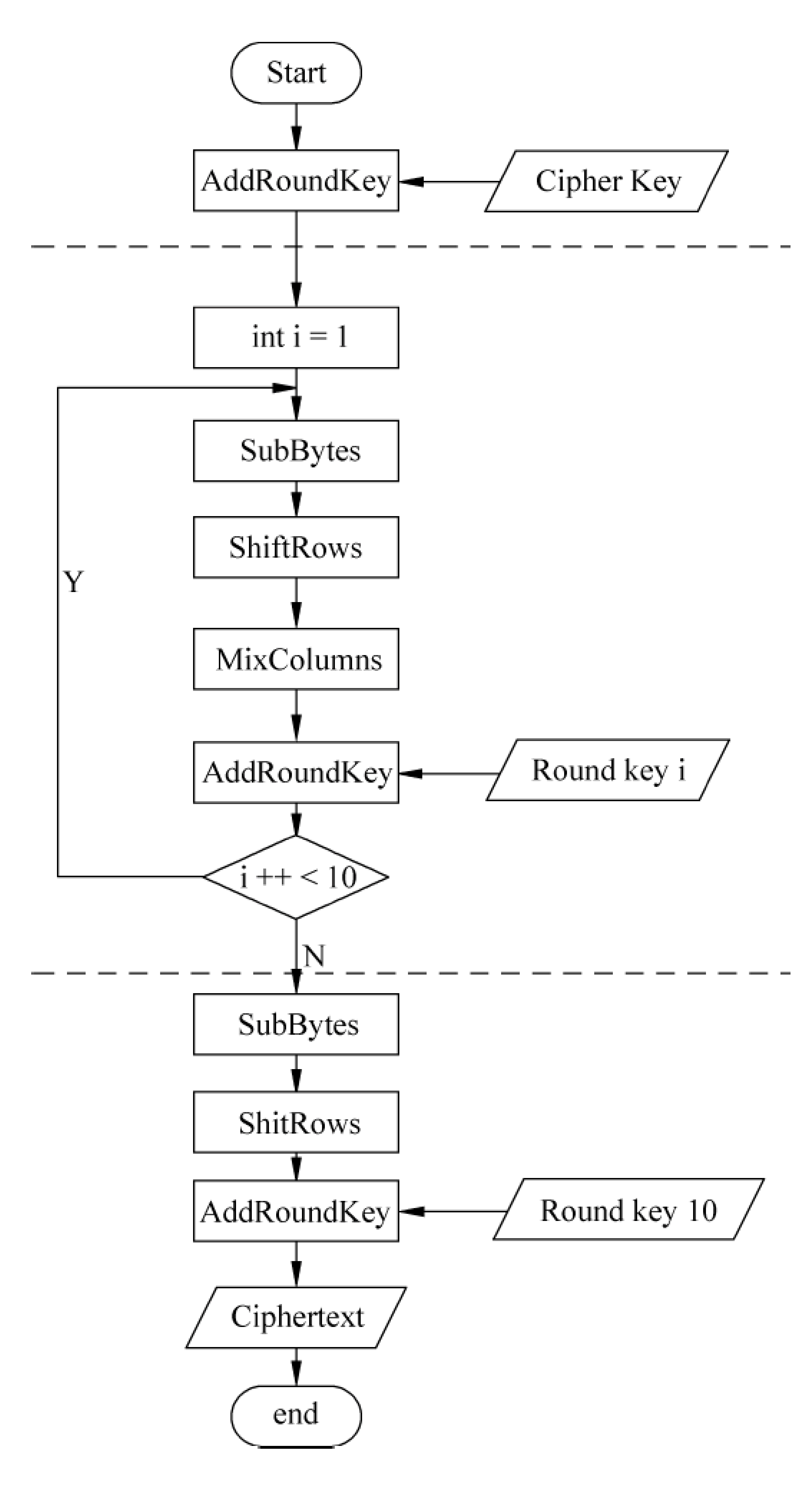


图 5-3　AES 加密算法结构

一是加密密钥与解密密钥是一对密钥，如果仅仅知道密码算法和加密密钥（即公开密钥），而要确定解密密钥（私钥），在计算上是不可行的。

二是大多数公钥密码算法的加密密钥和解密密钥具有互换的性质，即两者是相对的。如 RSA 算法，密钥对中的一个用于加密，另外一个就可用于解密。

非对称密码体制很好地解决了密钥管理问题。在对称密码体制中，密钥管理的工作量很大。例如，一个由 n 个人组成的团体，每个用户必须对和其他 $n-1$ 个用户之间的密钥保密，每个用户的密钥保管量将达到 $n(n-1)/2$ 个，当 n 比较大时候，这是一个困难的管理问题，因为密钥需要保密。但是在非对称密码体制下，这一问题要容易得多。

在非对称密码体制下，每个用户公布一个公开密钥，而保密自己的唯一的秘密密钥。任何一个用户 A 为了给用户 B 发送消息 M，只需要查找用户 B 的公开密钥，然后利用 B 的公开密钥将消息 M 加密，将密文通过不安全信道发送给用户 B。用户 B 收到密文后，利用自己的秘密密钥解密而获得明文。其他的用户，即使截获到密文，也不能解密获得明文。

显然这里存在一个问题，用户 A 如何得到用户 B 的公开密钥，即用户 A 如何相信某一个公开密钥是用户 B 的公开密钥？这一问题在公钥密码发展的初期阶段，人们以为可以像查电话号码一样在一个公开的密钥簿上查找公开密钥即可，但是，密码学家很快地意识到这可能带来欺诈，在实际中是不可行的。

针对这一可能的安全隐患，密码学家提出了公钥基础设施的概念，即建立一套公钥基础设施，由可信任的第三方，给每个用户颁发公钥证书，将每个用户的身份和他的公开密钥捆绑起来，并由可信第三方进行数字签名，这样任何一个用户都可以验证另外一个用户的公开密钥是否可以信任。

在非对称密码体制中，用户密钥管理的工作量被大大减少。非对称密码体制的保密通信模式，尤其适合于现在互联网时代的通信要求。因为在互联网这样的开放网络环境中，位于异地的通信双方不可能像传统的对称密钥密码体制所要求的那样，互相见面协商好会话密钥。非对称密码体制，使得双方即使没有互相见面，也可利用公钥证书确认对方的身份，利用对方的公开密钥实现秘密通信。

1）非对称密码体制的原理

非对称密码体制的正确运行需要满足以下四个要求：

（1）通信的参与方 B 容易通过计算产生一对密钥，即公开密钥 PK_B 和私有密钥 SK_B。

（2）信息的发送方 A 希望以秘密方式发送信息 M 给 B，在知道 B 的公开密钥 PK_B 时，可以通过加密算法生成密文：

$$C = E_{PK_B}(M)$$

（3）接收方 B 使用私有密钥容易通过解密算法对密文进行解密，以恢复原来的明文。

$$M = D_{SK_B}(C) = D_{SK_B}(E_{PK_B}(M))$$

（4）从公开密钥 PK_B 推出私有密钥 SK_B，在计算上是不可行的。

事实上，非对称密码体制的思想与单向陷门函数有关，它们都是依赖于数学上的困难问题而设计的，这一点与分组密码有着本质的区别。在非对称密码学思想被提出后，相继有几个具体的非对称密码算法被提出，其安全性依赖于不同的计算困难性问题。例如，最著名的 RSA 非对称密码体制的安全性依赖于大整数分解的困难性；ElGamal 非对称密码体制及其变种的安全性依赖于离散对数问题的困

难性。

非对称密码体制要满足上述要求，实际上就是需要设计一个单向陷门函数(One-way Trapdoor Function)。非对称密码体制中的公钥用于单向陷门函数的正向(加密)计算，而私钥用于反向(解密)计算。

定义 5.1 单向陷门函数，是指满足下列条件的函数 $f:D\to V$：

(1) 对于任意给定的 $x\in D$，计算 $y=f(x)$是容易的。

(2) 对于几乎所有任意给定 $y\in V$，计算 $x\in D$ 使得 $y=f(x)$，在计算上是困难的，即计算 $x=f^{-1}(y)$是困难的。这里所谓困难是指在有意义的时间要求之内计算是不可行的。

(3) 存在陷门信息 t，当已知 t 时，对给定的任何 $y\in V$，若相应的 x 存在，则计算 x 使 $y=f(x)$是容易的。

说明：

(1) 仅满足上面定义中的(1)、(2)两条的称为单向函数；第(3)条称为陷门性质，其中的 t 称为陷门信息(trapdoor information)。

(2) 这里的陷门信息 t 保密，可作为解密密钥，此时 t 称为秘密密钥(Private key)，即为 Sk。由于加密函数是公开的，任何人都可以将信息 x 加密成 $y=f(x)$；只有拥有陷门信息 t 的人，即拥有 Sk 才能解密出信息 $x=f^{-1}(y)$。

(3) 单向陷门函数的第(2)条性质可以确保安全性，即窃听者由截获的密文 $y=f(x)$推测明文 x 是不可行的。

2) RSA 算法

目前，RSA 是最著名的、也是应用最广泛的非对称密码体制，它是由 Rivest，Shamir 和 Adleman 三位密码学家在 1978 年提出的，算法用他们的名字命名。RSA 是基于 Diffie 和 Hellman 所提出的非对称密码学思想，是首个非对称密码思想的具体实现。该算法的数学基础是初等数论中的 Euler 定理，算法的安全性建立在大整数因子分解困难性的基础上。

算法 5.1 给出了 RSA 密码体制的具体描述，这是 RSA 算法最原始的形式。经过对 RSA 长期的深入研究，密码学家们已经提出了更好的、更安全的使用形式(非对称加密填充 OAEP 形式)。

算法 5.1 RSA 密码体制。

密钥的生成用户 Alice 为了生成 RSA 密钥执行以下步骤：

随机选择两个大素数 p 和 q，计算 $N=pq$ 和 $\phi(N)=(p-1)(q-1)$；随机选择整数 e，$1<e<\phi(N)$，满足 $\gcd(e,\phi(N))=1$，并计算整数 d 满足 $ed\equiv 1(\mathrm{mod}\,\phi(N))$；

用户 Alice 公开她的公钥 $PK=(N,e)$，安全地销毁 p、q 和 $\phi(N)$，并秘密保留她的私钥 $\mathrm{Sk}=(d)$。

加密用户 Bob 为了将消息 $m \in Z_N$ 秘密地发送给 Alice，Bob 首先获得 Alice 的公开密钥 PK＝(N,e)，计算密文 $c \in Z_N, c \leftarrow m^e \pmod N$。将密文 c 通过不安全信道发送给 Alice。

由 RSA 密码体制的定义可以看到，其中加密和解密运算都是在模 N 的整数环 Z_N 中进行的，主要用到模 N 的乘法运算。由于 N 是大合数，整数环 Z_N 中在乘法运算下可逆的元素组成的乘法运算群 $Z_N^* = \{a \mid 1 \leqslant a \leqslant N-1, \gcd(a,N)=1\}$，它的阶数，即 Z_N^* 中元素的个数记为 $|Z_N^*| = \varphi(N) = (p-1)(q-1)$。

下面证明 RSA 算法的正确性，也就是算法中的加密和解密互为逆运算。

由模运算的定义，算法 5.1 中的同余式 $ed \equiv 1 (\bmod \varphi(N))$ 意味着存在某个整数 k，使得 $ed = 1 + k\varphi(N)$。于是，设消息 $m \in Z_N^*$，Alice 的解密过程得到的数是

$$c^d \bmod N \equiv m^{ed} \bmod N \equiv m^{1+k\varphi(N)} \bmod N \equiv m$$

所以，解密正确。

下面给出一个简单的例子。

【例 5.1】 RSA 算法实例。

(1) 设通信方 Alice 选择两个素数 $p=7$ 和 $q=13$，计算 $N=7\times 13=91$ 以及 $\varphi(N)=6\times 12=72$。

(2) 选择一个随机整数 $e=5$，满足 $\gcd(5,72)=1$。

(3) 由欧拉辗转相除法，根据 e 和 $\varphi(N)$，Alice 计算得到 d：

$$72\times(-2)-5\times 29=1$$

即 $5\times 29=1(\bmod 72)$。于是，Alice 计算得到的 $d=29$ 作为她秘密的解密密钥，Alice 公开 $(N,e)=(91, 5)$ 作为 RSA 体制的公开密钥。

(4) 设 Bob 欲秘密发送明文 $m=3$ 给 Alice，Bob 计算 $c=3^5=243\equiv 61(\bmod 91)$，即得到的密文是 61。

(5) 为了解密密文 61，Alice 利用自己的秘密钥 d 计算 $61^{29}\equiv 3(\bmod 91)$，即得到了明文 $m=3$。

需要说明的是，这是一个示例性的例子，实际中所使用的 N 的大小在 1025 比特以上，甚至达到 2048 比特。即 N 是一个大整数，只有这样才能保证其因子分解的困难性，即必须保证在模 N 的整数环 Z_N 中求逆运算的困难性，才能使得从公开密钥 (N,e) 不能推出秘密钥 d。

在上面的 RSA 非对称密码体制中，使用了单向陷门函数的概念，这里的单向陷门函数是

$$c=f(m)\equiv m^e (\bmod N)$$

一方面，如果已知公钥 PK＝(N,e) 及消息 m，可以很容易地计算得到密文 c。另一方面，在不知道秘密钥 d 时，从密文 c 求出消息 m，这是一个困难问题。但是

如果知道陷门信息 $t=\mathrm{Sk}=\{d\}$，公钥 $\mathrm{PK}=(N,e)$ 及密文 c，则可以求解消息 m。大整数分解的困难性保证了从公开钥 (N,e) 不能推出秘密钥 d。

3）非对称密码体制的应用

非对称密码体制的特点就是每个用户有两个密钥，其中一个是保密的，而另一个则是公开的。用公钥加密得到的密文可以用相对应的私钥进行解密；反之，用私钥加密得到的密文也可以用相对应的公钥解密。非对称密码系统的这些特点使其可以在信息安全的三个方面发挥作用。

（1）信息机密性的实现。

信息的发送方用期望的接收方的公钥加密报文，则只有该接收者可以用自己的私钥解密该信息，而通信链路上的其他人即使非法截获了密文信息，也无法获得明文。

（2）不可否认性的实现。

为了在网络通信中实现发送消息的不可否认性，即发送方不能否认曾经发送了某个消息，可以借助数字签名技术来实现。而非对称密码体制是实现数字签名的基础。事实上，数字签名就是发送方首先对于要发送的消息生成消息摘要，并用自己的私钥对消息摘要进行加密而得到的密文；消息的接收方可以用发送方的公钥解密数字签名还原出消息摘要，并与重新计算得出的消息摘要进行比对，以验证消息是否被篡改。通过这一过程，使得消息的发送方不能否认自己曾经发送过该消息，因为接收方用发送方的公钥从数字签名中还原出了消息摘要，说明一定是由发送方私钥加密得到的该数字签名。

（3）密钥交换。

前面已经讨论过，对称密码体制的优点是加解密的效率高，而在开放网络环境下存在的一个应用障碍则是密钥分发困难，即在信息的接收方和发送方之间如何交换密钥的问题。在非对称密码体制的支持下，可以构建数字信封，实现发送方和接收方之间的密钥交换。数字信封技术如图 5-4 所示，通过非对称密码体制实现了通信双方的对称密钥的安全共享。

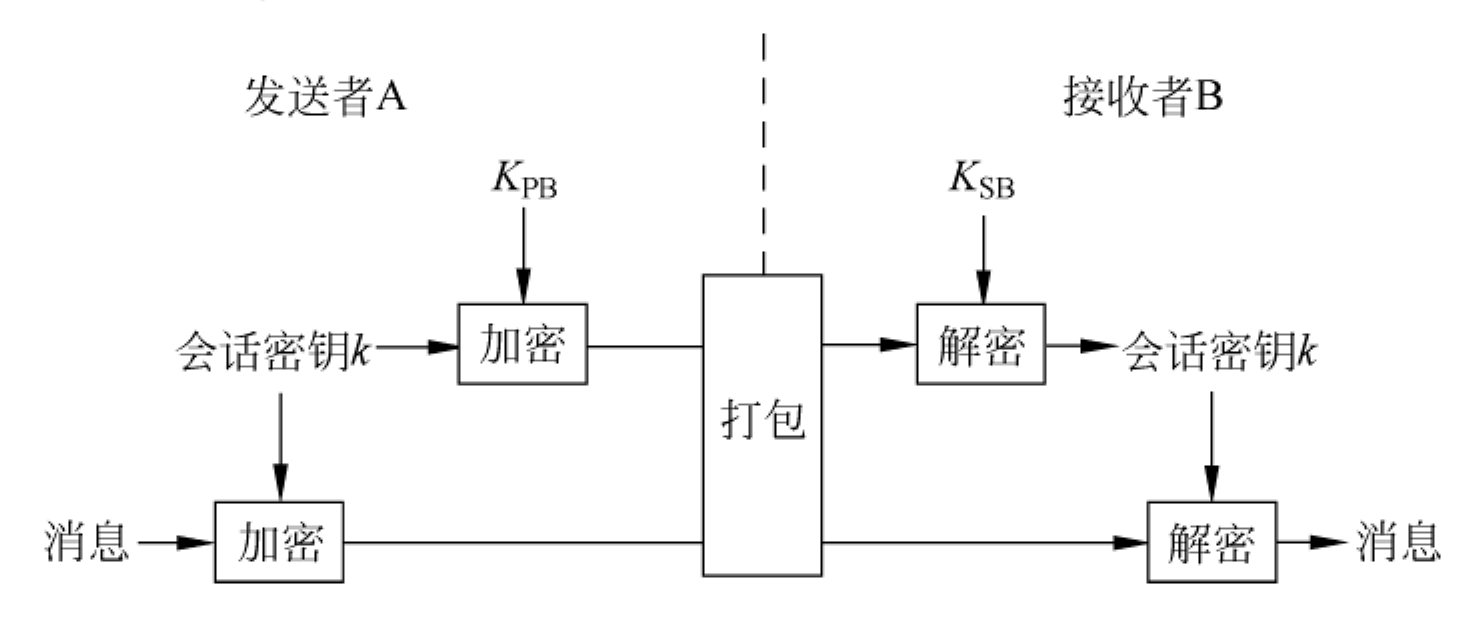


图 5-4　数字信封示意图

5.1.2 信息的完整性保障技术

在信息系统中,数字签名与身份认证技术是信息完整性和不可否认性的重要保障,是公钥密码体制的重要应用。信息的发送方可以对电子文档生成数字签名,信息的接收方则在收到文档及其数字签名后,可以验证数字签名的真实性。身份认证则是基于数字签名技术为网络世界中实体的身份提供可验证性。

1. 数字签名原理

在非对称密码系统中,由于安全性要求从公开密钥不能推算出私有密钥,所以公布公开密钥并不会威胁到私有密钥的安全;公开密钥是可以公开传播的,而私有密钥一定是由个人秘密持有的。因此,如果某人用其私有密钥对消息进行加密,密文如果能够使用他的公开密钥进行解密而得到明文,就可以肯定该消息的加密是某人用私钥完成的。因为其他的公开密钥是不可能正确解密该密文消息的,而其他人也不可能持有该私有密钥而计算出该加密消息。可以用图 5-5 表示非对称密码体制用于数字签名的原理。

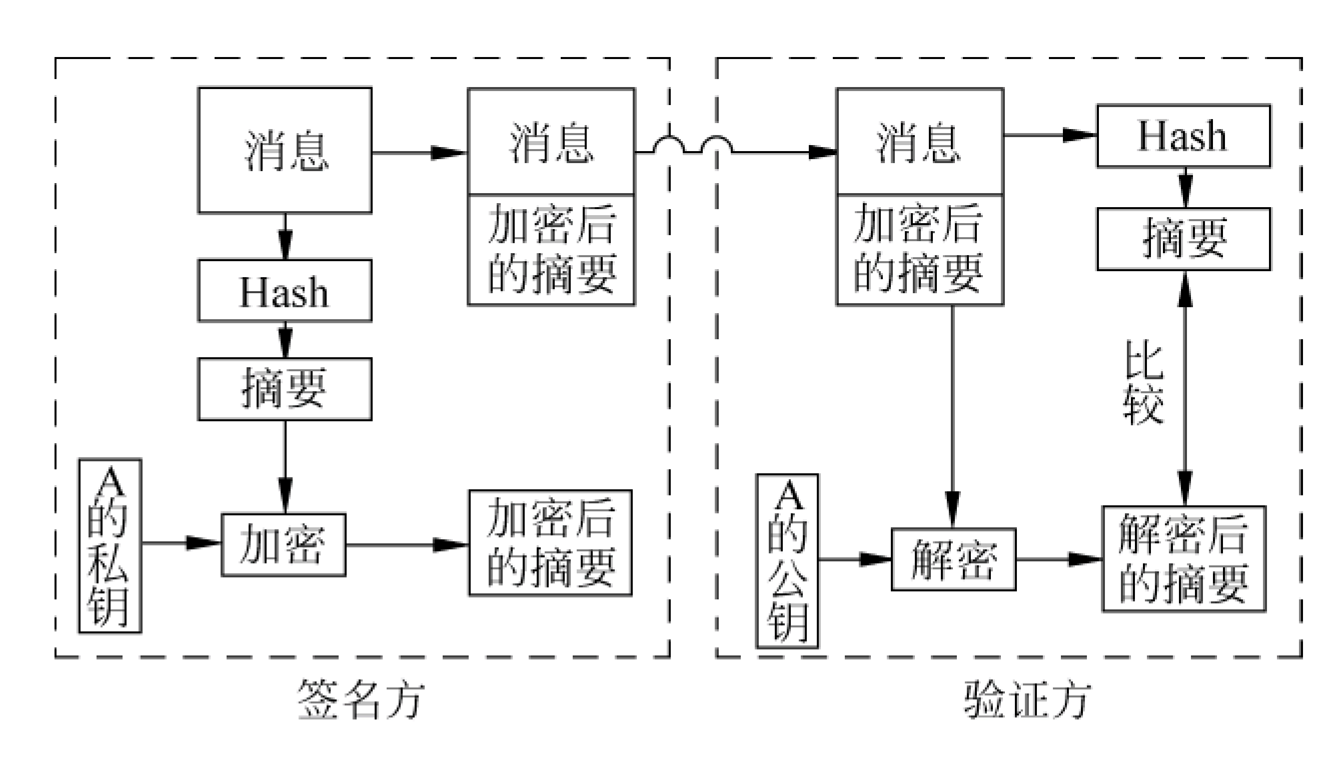


图 5-5 非对称密码体制用于数字签名的原理图

一个基于公钥密码学的数字签名方案被定义为一个算法三元组(Gen,Sig,Ver),方案中有两方参与者:签名者 Signer 与验证者 Verifier。

- 密钥生成算法 Gen:它是一个概率多项式时间算法,由系统或者由签名者执行,算法的输入为 1^k,成为系统安全参数,输出为密钥对(Pk,Sk),其中 Pk 为签名者的公开密钥,Sk 为签名者的秘密密钥;即:Gen(1^k)→(Pk,Sk);
- 签名生成算法 Sig:该算法是一个概率多项式时间的算法,由签名者执行,该算法以秘密密钥 Sk 和待签名的消息 $m\in\{0,1\}^k$ 为输入,输出为串 s,此时 s 称为签名者用秘密密钥 Sk 对消息 m 进行的签名,即:Sig(Sk,m)→s;
- 签名验证算法 Ver:它是一个确定性算法,由验证者执行,该算法以签名者

的公开密钥 Pk 及签名消息对(m,s)为输入，输出 0 或 1，即：Ver$(\mathrm{Pk},m,s)\to\{0,1\}$如果 $s\in\mathrm{Sig}(m)$，则输出 1 说明签名有效；反之输出 0，说明签名无效。

首先由可信任的第三方采用密钥生成算法为签名者生成密钥对(Pk，Sk)，将签名者的公开密钥 Pk 公开，将秘密密钥 Sk 由签名者秘密持有。当用户需要对某一消息 m 签名时，其采用签名算法 Sig 以自己的私钥 Sk 和 m 为输入得到消息 m 的签名 $s=\mathrm{Sig}(\mathrm{Sk},m)$。签名生成后，签名者将签名消息对$(m,s)$提交给验证者；验证者采用验证算法 Ver，以签名者公开密钥和消息签名对作为输入，验证签名是否有效，即：Ver$(\mathrm{Pk},m,s)\to\{0,1\}$。这一验证可以在事后任一个时间进行。一个经过验证确认是有效的签名，签名者必须对这一签名负责，因为假设签名方案是安全的，则只有签名者拥有签名的秘密密钥，因而只有签名者才能生成某消息的有效签名。

2. 哈希函数

定义 5.2 哈希(Hash)函数是一个输入为任意长的二元串，输出为固定长度的二元串的函数。一般用 $H(\cdot)$表示哈希函数，若输出是长度为 l 的二元串，则哈希函数表示为

$$H(\cdot):\{0,1\}^*\to\{0,1\}^l$$

其中$\{0,1\}^*$表示所有任意有限长的二元串的全体集合，$\{0,1\}^l$表示所有长度为 l 的二元串的集合。若消息 $M\in\{0,1\}^*$，则 $H(M)\in\{0,1\}^l$。哈希函数又称为散列函数或杂凑函数。

哈希函数的作用是将任意长度的二进制消息(文件)压缩成固定 l 比特长度的二进制串，目前密码学中使用的哈希函数的输出长度 l 一般取 128、160、192、256、320、385、512 比特等等，通常为 32 的整数倍。

定义 5.3 哈希函数 $H(\cdot):\{0,1\}^*\to\{0,1\}^l$ 称为具有单向性，是指

(1) 任意给定 $M\in\{0,1\}^*$，可以很容易(多项式时间内)计算出消息摘要 $H(M)\in\{0,1\}^l$；

(2) 任意给定 $H(M)\in\{0,1\}^l$，求出 $M\in\{0,1\}^*$ 是计算上困难的，即多项式时间内不可解。

通俗地讲，哈希函数的单向性是指任意给定 $M\in\{0,1\}^*$，可以很容易计算出消息摘要 $H(M)\in\{0,1\}^l$，反之，给定哈希函数值要推算出消息 M，则是难以计算的。任意给定 $H(M)\in\{0,1\}^l$，求出 $M\in\{0,1\}^*$ 是计算上困难的这一性质也称为哈希函数 $H(\cdot):\{0,1\}^*\to\{0,1\}^l$ 是抗原像的(Preimage Resistant)。

如果将消息摘要作为消息的数字指纹，仅仅要求哈希函数具有单向性是不够的，还要求哈希函数必须具有抗碰撞性。下面给出抗碰撞性的定义。

定义 5.4 哈希函数 $H(\cdot):\{0,1\}^*\to\{0,1\}^l$ 称为具有抗第二原像性(Second Preimage Resistant),是指任意给定 $M\in\{0,1\}^*$ 及其信息摘要 $H(M)$,求出 $M'\in\{0,1\}^*$ 且 $M'\neq M$,使得 $H(M')=H(M)$是困难的。

显然,哈希函数的抗第二原像性使得消息 M 的信息摘要 $H(M)\in\{0,1\}^l$ 基本上可以作为消息 $M\in\{0,1\}^*$ 的标识符。但是,会不会任意两个不同的消息产生相同的消息摘要? 这是可能的。因此,有以下更进一步的定义。

定义 5.5 哈希函数 $H(\cdot):\{0,1\}^*\to\{0,1\}^l$ 称为具有抗碰撞性(Collision Resistant),是指求出任意 $M,M'\in\{0,1\}^*$,且 $M'\neq M$,使得 $H(M')=H(M)$是困难的。

显然,哈希函数 $H(\cdot):\{0,1\}^*\to\{0,1\}^l$ 的抗碰撞性使得消息摘要 $H(M)\in\{0,1\}^l$ 可以作为消息 $M\in\{0,1\}^*$ 的标识符,这是因为,求得具有相同的消息摘要的两个不同消息是困难的。

由上面的四个定义可以知道,哈希函数应该具有单向性、抗原像性、抗第二原像性以及抗碰撞性。具有这些性质的哈希函数才能够应用于数字签名技术中,实现消息的完整性检验。

如何将输入的任意有限长的二元串 $M\in\{0,1\}^*$ 压缩成固定长度的输出,是设计哈希函数 $H(\cdot):\{0,1\}^*\to\{0,1\}^l$ 面临的首要问题。目前哈希函数一般都采用 Merkle-Damgard 迭代结构实现,由 Merkle 提出的迭代哈希函数一般结构如图 5-6 所示,这也是目前大多数哈希函数(MD5、SHA-1、RIPEMD)的通用结构。其中,IV 称为初始向量,CV 称为链接变量,Y_i 是第 $i+1$ 个输入消息分组,f 称为压缩函数,L 为输入的分组数,l 为哈希函数的输出长度,b 为输入分组长度。

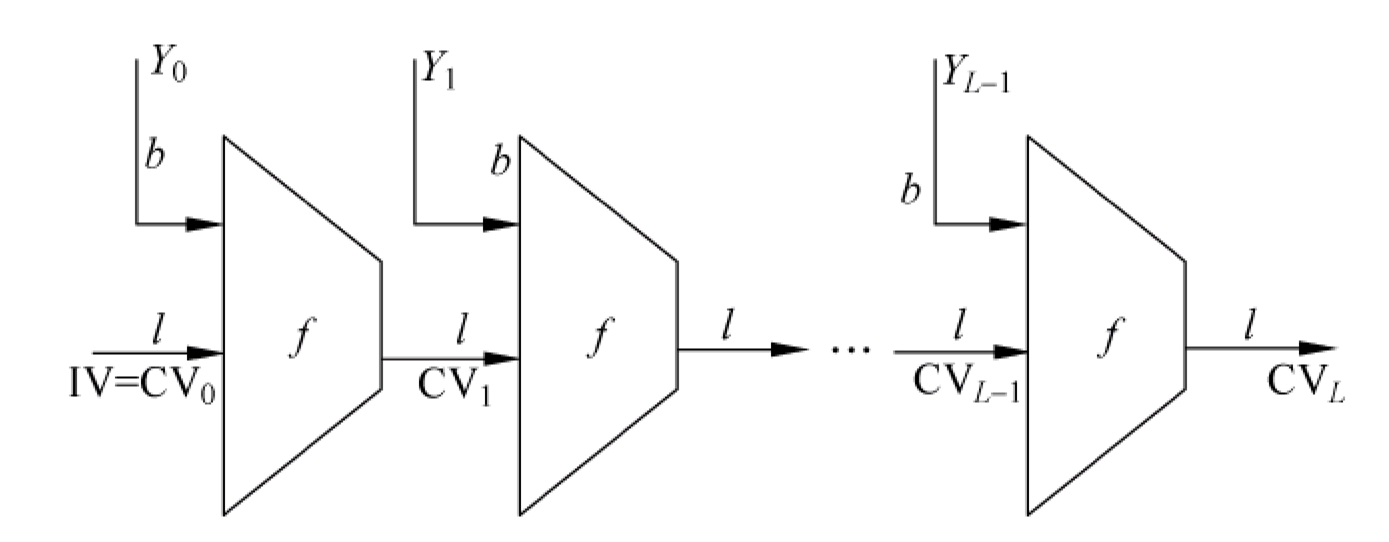


图 5-6 迭代哈希函数一般结构

该方法涉及两个设计步骤,其中包括消息填充方法和一个压缩函数。

首先,将有限长的输入二元串 $M\in\{0,1\}^*$ 填充为长度恰好为 b 比特的若干个数据块(Block)。一个典型填充方法是在输入消息二元串 M 后面添加一个比特 1,然后填充足够多的比特 0,再添加 M 的长度 $|M|$ 的二进制表示,使填充后的消息的长度为 b 的 L 整数倍,设填充后消息表示为 $Y_0||Y_1||\cdots||Y_{L-1}$。

其次,设计一个压缩(Compress)函数 $F(\cdot):\{0,1\}^{b+l}\to\{0,1\}^l$,$b>l$。注意,

这里压缩函数 $F(\cdot):\{0,1\}^{b+l}\rightarrow\{0,1\}^l$ 的输入是固定长度 b 比特信息(目前,大多数情况取 $b=512$bit),这里 $b>l$ 表示这个函数是一个消息压缩的过程。

取一个初始值 $\mathrm{IV}\in\{0,1\}^l$,然后计算:$\mathrm{CV}_0\leftarrow\mathrm{IV}$,$\mathrm{CV}_1\leftarrow F(Y_0||\mathrm{CV}_0)$,$\mathrm{CV}_2\leftarrow F(Y_1||\mathrm{CV}_1)$,……,$\mathrm{CV}_L\leftarrow F(Y_{L-1}||\mathrm{CV}_{L-1})$。最后,$\mathrm{CV}_L$ 的值即被作为哈希函数的输出,如图 5-6 所示。

最著名的哈希算法有 MD5、SHA 以及 RIPEMD-160 等,这里以安全哈希函数 SHA 为例进行介绍。安全哈希算法(SHA)是由美国国家标准和技术协会(NIST)提出的,于 1993 年作为美国联邦消息处理标准(FIPS PUB 180)公布。1995 年 NIST 发布了它的修订版(FIPS 180-1),通常称为 SHA-1。

3. RSA 数字签名体制

在 Diffie 和 Hellman 于 1976 年首次提出数字签名概念后,RSA 签名体制是第一个数字签名体制,它由 Rivest、Shamir 和 Adleman 三人共同提出。RSA 签名体制的详细说明见算法 5.2。

算法 5.2　RSA 签名体制。

密钥建立

密钥建立过程和 RSA 密码系统的密钥建立过程(见算法 5.1)相同。经过密钥建立过程,用户 Alice 的公钥为(N,e),其中 $N=pq$,p 和 q 是两个长度差不多的大素数,e 是满足 $\gcd(e,\varphi(N))=1$ 的整数。Alice 的私钥为 d,满足 $ed=1 \bmod (\varphi(N))$。

签名生成

为了生成消息 $m\in Z_N^*$ 的签名。Alice 计算 $s=\mathrm{Sign}_d(m)\leftarrow m^d (\bmod N)$,即得到消息签名对$(m,s)$。

签名验证

设 Bob 是验证者,他知道公钥(N,e)属于 Alice。给定一个消息-签名对(m,s),Bob 的验证过程为测试 $m\equiv s^e (\bmod N)$,如果成立,则 $\mathrm{Verify}_{(N,e)}(m,s)=\mathrm{True}$。

很容易看出,RSA 数字签名过程与 RSA 加密和解密过程非常类似,唯一不同的是,现在 Alice 首先用她的私钥进行“加密”,而 Bob(或任何人)再用 Alice 的公钥进行“解密”。

需要说明的是,算法 5.2 直接用于实际的数字签名很不安全,任何知道 Alice 公钥 e 的人都可以容易地伪造签名。例如,Bob 可以选取一个随机数 $s\in Z_N^*$,并计算

$$m\leftarrow s^e \bmod N \tag{5.2}$$

然后声称该(m,s)是由 Alice 签名的消息签名对。

显然,这样伪造的“消息”-签名对(m,s),完全可以通过 RSA 的验证算法,即验证结果为 True。对于基本的 RSA 签名体制的攻击,除了上述伪造攻击之外,RSA

算法中的乘法性质也为攻击提供了一个简单的方法，使得攻击者可以从已知的消息-签名对伪造新的消息-签名对。例如从现有的消息-签名对(m_1, s_1)和(m_2, s_2)可以伪造出一个新的消息-签名对$(m_1 m_2, s_1 s_2)$。显然，伪造的新消息-签名对可以通过 RSA 的验证算法，关于这一点读者可自行证明。

上述两种伪造签名的攻击方法属于存在性伪造。所谓的存在性伪造，是指伪造出的消息-签名对中的消息是随机信息，不具有实际意义。显然，通过式(5.2)得到的或通过将两个已有签名的消息相乘所生成的新消息 m 是随机的。

可以通过为 m 增加一些可识别的冗余信息，使之变得不随机或“是有意义的”来抗击这种存在性伪造。为消息增加可识别信息有两种，其中的简单方法是使消息本身包含可识别的部分。例如 $m=M||I$，其中 M 是真正要签名的消息；I 为一个可识别的串，比如签名者的身份。

消息增加可识别信息的另一个常用方法是利用密码学哈希函数(参见 5.1.2 节)对该消息进行“散列”。在数字签名中，对消息进行散列的另一个作用是将任意长度的消息压缩成固定长度的消息指纹。具体的算法改为：为了生成消息 $m \in Z_N^*$ 的签名，Alice 生成 $s \leftarrow h(m)^d \pmod N$，即得到消息-签名对$(m, s)$。收到消息-签名对$(m, s)$，Bob 的验证过程为 $\text{Verify}_{(N,e)}(m, s) = \text{True}$(若 $h(m) \equiv s^e \pmod N$)。

在这个改进的 RSA 签名算法中使用对消息进行散列的方法，使得对签名的存在性伪造不能奏效，故该算法才能真正用于数字签名应用中。因为攻击者无法知道所用密码哈希函数下 $H(m)$的原像，那么像式(5.2)那样从一个随机选定的 s 计算 m 就不可能完成，故无法给出伪造的消息-签名对。

5.1.3 消息认证技术

消息认证是使消息的接收者能够检验收到的消息是否真实的认证方法。消息认证的目的有两个，其一是对消息源的认证，即验证消息的来源是真实的；其二是对消息的认证，即验证信息在传送过程中未被篡改。

有两类方法用来进行消息认证：

消息认证码(Message Authentication Code，MAC)：是以消息和密钥作为输入的公开函数，可以生成定长的输出。该方法需要在信息的发送方和接收方之间共享密钥。

哈希函数：是不带密钥的公开函数，它将任意长度的输入消息映射为固定长度的输出值。哈希函数与数字签名算法相结合，提供对于消息的完整性检验。

其中，最常用的基于密钥哈希函数的 MAC 的形式为：

$$\text{MAC} = H(k \parallel M) \tag{5.3}$$

为了提供一个消息 M 的认证性，发送者通过式(5.3)计算消息的 MAC，其中，

k 为发送者和接收者的共享密钥，“‖”表示比特串的链接。在接受方与发送方共享密钥情况下，将密钥作为哈希函数的一部分输入，另一部分输入为需要认证的消息。

根据 5.1.2 节中讨论的哈希函数的性质，可以假设，为了应用哈希函数生成一个关于密钥 k 和消息 M 的有效 MAC，该主体必须拥有正确的密钥和正确的消息。接收者利用接收的消息 M 及与发送者共享的密钥 k 重新计算出 MAC，并与所收到的 MAC 比较是否一致。如果一致，就可以相信该消息来自所声称的发送者，并在传输中未被篡改。

这种使用哈希函数构造的 MAC，称为 HMAC（用哈希函数构造的 MAC）。为谨慎起见，HMAC 通常通过式(5.4)计算

$$\mathrm{HMAC} = H\ (k \parallel M \parallel k) \tag{5.4}$$

也就是说，将密钥作为认证消息的前缀和后缀，这是为了阻止攻击者利用某些哈希函数的“轮函数迭代”结构。如果不用密钥保护消息的两端，而采用式(5.3)的形式，如果攻击者已知哈希函数具有的这种结构，使得攻击者不必知道密钥 k 就可以选择一些数据用作消息前缀或后缀来修改消息，使得其可以通过认证码的认证。

5.1.4　身份认证技术

认证是一个实体向另一个实体证明某种声称的属性的过程。认证包括数据源认证（Data-origin Authentication）和实体身份认证（Entity Authentication）。例如，前者是一个主体，声称拥有某种合法权利，可以进入后者的系统或者使用后者的服务，通过认证，后者确认其确实拥有这种权利。一般地，认证至少涉及两个独立的通信实体，一个认证过程也就是一个认证协议。

1. 数据源认证

数据源认证，与数据完整性密切相关。早期的密码学和信息安全教程认为数据源认证与消息认证没有本质区别，这种观点是基于以下的考虑：使用被恶意修改过的信息和使用来源不明的消息具有相同的风险。然而，数据源认证和数据完整性在概念上差别很大，它们在很多方面都有明显区分。

首先，数据源认证一定涉及一个通信过程。在这种安全服务中，消息接收者对于消息是否来源于其所声称的消息源进行验证；而数据完整性不一定涉及通信过程，这种安全服务可以应用于存储中的数据；其次，数据源认证一定涉及对消息源的识别，而数据完整性服务则不一定涉及该过程；再次，也是最重要的一点，数据源认证一定涉及消息的新鲜性（Freshness）确认，而数据完整性则不必认证消息的新鲜性，因为老的数据也可能需要有数据完整性。为了获得数据源认证服务，消息的接收者应该验证该消息是否是新近发送的（也就是说，消息的发送和接收之间的

时间间隔应该足够小)。新近发送的消息被称为新鲜的消息,接收者要求消息的新鲜性是符合常识的,新鲜消息就意味着在通信双方之间的通信是一个良好的状态。例如,针对 Needham-Schroeder 对称密钥认证协议的攻击 Denning 和 Sacco 攻击,该攻击中的一条重放的旧消息就具有有效的数据完整性而没有有效的认证性,这种类型的认证失败称为缺失消息源活现性的有效数据完整性。

归纳起来,数据源认证的特征可以总结为:数据源认证包含从发送者到接收者的消息传输过程,接收者在接收时会验证消息,接收者执行消息验证的目的在于确认消息发送者的身份;确认在原消息离开消息发送者之后的数据完整性以及确认消息传输的“活现性”。

2. 身份认证

通信实体可以是一个人、一个程序、一个客户机或一个服务器。需要验证身份的实体称为原告;试图证明原告身份的一方称为验证者。在身份认证中,验证者要考虑与认定的通信方的通信真实性,为此,证明消息新鲜性或主体活现性的机制就成为身份认证协议中最基本的一个组成部分,它主要是通过询问-应答机制(Challenge-response Mechanisms)实现的。

基于密码技术实现身份认证可以采用对称密码技术也可以采用非对称密码技术,例如,X. 509 认证技术就是基于非对称密码技术的,Kerberos 认证技术是基于对称密码技术的。如果机制中使用的是对称密码技术,则身份认证的双方 Alice 和 Bob 必须共享某个密钥 KAB;如果机制用的是非对称密码技术,验证方 Bob 必须能够通过公钥证书框架知道被验证方 Alice 的公钥。下面,以 X. 509 认证为例说明身份认证技术。

3. X. 509 认证技术

基于 X. 509 证书的认证技术类似于 Kerberos 技术,它也依赖于共同信赖的第三方以实现认证。与 Kerberos 认证不同的是,X. 509 认证技术采用的是非对称密码体制(公钥制),在 X. 509 认证框架中可信赖第三方是指一个称为 CA (Certificate Authority)的认证机构。该认证机构负责核实用户的身份,并为用户签发数字公钥证书,同时对于证书提供管理。其中数字证书遵循 X. 509 标准中所规定的格式,称为 X. 509 证书。持有此证书的用户可以凭证书访问信任 CA 的服务器。

当用户向某一服务器发起服务请求时,服务器则要求用户必须提交数字证书。收到用户的数字证书后,服务器对证书进行验证,首先利用 CA 的公开密钥对证书中的 CA 签名进行解密,以获得证书文本信息的散列码;然后用与 CA 相同的散列算法对证书文本信息进行 hash,得到一个证书信息的散列码;将此散列码与对签名解密所得到的散列码两者进行比较,若相等则说明这一证书是由 CA 签发的,而

且是完整的、未被篡改的。这样，用户便通过了身份认证。服务器从该证书的信息部分取出该用户的公钥，如果以后需要向用户传送数据时，以此公钥进行加密。这样，只有该用户可以解密服务器的加密信息，从而保证了用户与服务器之间的通信机密性。

基于X.509证书的认证技术适用于开放式网络环境下的身份认证，该技术已被广泛接受，许多网络安全程序都使用X.509证书（如IPSec、SSL、SET、S/MIME等）。

X. 509证书的认证框架使用公钥密码学的技术识别通信方，根据要求的认证强度的不同，提供单向认证、双向认证、三向认证三种认证模式。

1）单向认证

这种认证方式适合于通信的一方向另一方证实自己的身份的情况，协议只需要一次通信。例如用户A向用户B发送一条消息，消息的内容如下：

$$A \rightarrow B: t_A \,||\, R_A \,||\, \mathrm{ID}_B \,||\, \mathrm{sgn\ Data} \,||\, E_{K_{\mathrm{PB}}}[K_{\mathrm{AB}}] \,||\, \mathrm{signature}_A$$

t_A 表示时间戳，一般由两个日期组成：消息的生成时间和期满时间，时间戳用来防止消息传递的延迟及抗重放攻击。R_A 是一次性随机数，在一个有效期内是唯一的；ID_B 是A希望通信的B的身份信息；sgn Data是认证消息携带的数据信息；K_{AB}是A为以后的通信随机选择的会话密钥，并用B的公钥 K_{PB}对会话密钥进行了加密，最后附上A用自己的私钥对于前面所有内容的生成的数字签名 $\mathrm{signature}_A$。

B在收到上述消息后，将通过向CA查询A的证书，用A的公钥验证消息的签名 $\mathrm{signature}_A$，如果验证通过，则一方面B可以确认消息的发送方是A；另一方面B也就可以确认消息 $t_A||R_A||\mathrm{ID}_B||\mathrm{sgn\ Data}||E_{K_{\mathrm{PB}}}[K_{\mathrm{AB}}]$的完整性。随后，B可以核对消息中的 ID_B 和自己身份的一致性，接着用自己的私钥解密 $E_{K_{\mathrm{PB}}}[K_{\mathrm{AB}}]$从而获得会话密钥 K_{AB}。至此，B就认证了A的身份，同时取得了双方通信的会话密钥，可以用该密钥实施双方之间的保密通信。

2）双向认证

双向认证协议是由两次通信构成的，可以实现通信双方的互相鉴别，其过程如下：

（1）$A \rightarrow B: t_A||R_A||\mathrm{ID}_B||\mathrm{sgn\ Data}||E_{K_{\mathrm{PB}}}[K_{\mathrm{AB}}]||\mathrm{signature}_A$

（2）$B \rightarrow A: t_B||R_B||\mathrm{ID}_A||R_A||\mathrm{sgn\ Data}||E_{K_{\mathrm{PA}}}[K_{\mathrm{AB}}]||\mathrm{signature}_B$

这里，A发给B的消息内容以及B收到后所进行的验证操作均和单向认证相同，不再赘述。B在认证了A的身份之后，通过消息（2）向A证实自己的身份。首先B生成一个非重复的随机数 R_B，作用与 R_A 相同，在消息中包括A的身份 ID_A 以及从消息（1）中得到的随机数 R_A，以表明B正确接收了上一个来自于A的消息，同时B选定将用于加密给A数据所用的会话密钥，并用A的公钥进行加密，最

后用自己的私钥对于 $t_B||R_B||ID_A||R_A||$sgn Data$||E_{K_{PA}}[K_{AB}]$进行签名。

A 收到消息后首先需要获取 B 的证书，并验证证书的有效性。从 B 的证书中提取公钥，验证 B 的签名，同时检验消息的完整性。检查 A 自己是否是消息的接收者。验证时间戳 t_B 是否为当前时间。检查 R_B 是否被重放（可选）。至此，通信方 A 和 B 都认证了对方的身份，并得到了进一步保密通信的会话密钥。

3）三向认证

由于双向认证的最后，通信方 B 无法确认 A 是否正确接收了协议消息(2)，为此，可以采用更加完备的三向认证，其过程如下：

(1) $A \rightarrow B: t_A||R_A||ID_B||$sgn Data$||E_{K_{PB}}[K_{AB}]||$signature$_A$

(2) $B \rightarrow A: t_B||R_B||ID_A||R_A||$sgnData$||E_{K_{PA}}[K_{AB}]||$signature$_B$

(3) $A \rightarrow B: R_B||$signature$_A$

在三向认证的协议中，增加了第(3)条从 A 到 B 的消息，其中包含了来自于第(2)条消息中 B 所发送的一次性随机数 R_B，并且 A 用自己的私钥对其进行了签名。这样，对于收到的协议消息(2)进行了确认。

由于每个协议消息都包含了上一个协议消息携带的一次性随机数，每一端都可以通过检查返回的一次性随机数来探测重放攻击。所以三向认证可以不需要时钟同步，在不具备时钟同步条件时，可以采用这种方法。

4. 认证技术在 IPSec 中的应用

这里，以 IPSec 中的验证报头(Authentication Header)为例，说明消息认证和身份认证在 IPSec 的应用。

验证报头 AH 支持数据的完整性和 IP 包的验证。数据的完整性特征可以保证数据在传输中一旦被篡改，即可检测到。身份验证功能使得末端系统可验证用户或应用程序；同时，验证报头还能防止地址遭到欺骗攻击，并阻止重放攻击。

Authentication Header 包含如下字段：

Next Header(8 位)——识别这个报头之后紧跟的报头类型。

Payload Length(8 位)——有效载荷长度，Authentication Header 的长度(以 32 位字为单位)减 2。

Reserved(16 位)——保留作将来使用。

Sequence Number(32 位)——序列号，为一个单调递增的计数器值，用于抗重放攻击。

Authentication Data(可变长)——验证数据，也被称为完整性校验值(ICV)，必须是 32 位字的整数倍，为 MAC 码，目前的规范为 96 位。

在 Authentication Data 部分一般是消息验证码值，当前的规范支持两种规范：HMAC-MD5-96 和 HMAC-SHA-1-96。其中，前者使用包含 MD5 的 HMAC 算

法，后者使用包含 SHA-1 的 HMAC 算法，二者产生的 Authentication Data 长度均为 96 位。

ICV 是 AH 用来验证 IP 数据包的完整性所用的验证数据，是用 MAC 生成的。在生成过程中，必须使用双方共享的密钥。在进行通信之前，双方需要首先进行 SA 协商。SA 是一个在通信双方之间进行的关于参数选择的协议，这些参数中包括用来认证或者加密数据的密码算法、协商密码算法中使用的密钥、IPSec 协议(AH 或者 ESP)以及协议操作模式和生命期等。

在通信双方之间建立了 SA 之后，他们就有了所有用来计算他们交换的数据包的 ICV 的参数。ICV 的计算涉及整个 IP 头，然而有些域在传输过程中可能会改变，所以在计算 ICV 时需要将这些域设为 0。例如在 IPV5 头中的不变域包括版本、头长度、总长度、标识、协议、源地址、目标地址，以及数据(被封装的传输协议头和数据)；而可变域包括服务类型、标志、分段偏移量、TTL、头校验值。

除了 AH 之外，IPSec 中的 ESP (Encapsulating Security Payload，封装安全载荷)被用来提供保密性、数据来源认证(鉴别)、防重放攻击服务，以及通过防止数据流分析来提供有限的数据流加密保护。ESP 提供和 AH 类似的服务，但增加了两个额外的服务：数据保密和有限的数据流保密服务。保密服务是通过使用密码算法对 IP 数据报相关部分加密来实现。数据流保密则由隧道模式下的保密服务来提供。

ESP 中对数据报的加密算法都采用对称密钥体制。公钥密码算法涉及计算量非常大的大整数模指数运算，且大整数的规模超过 300 位十进制数字。而对称密码算法主要使用初级操作(异或、逐位与、位循环等)，以软件或是硬件方式执行都非常有效。所以相对公钥密码系统而言，对称密钥系统的加、解密效率要高得多。ESP 通过在 IP 层对数据包进行加密来提供保密性，它支持各种对称的加密算法。对于 IPSec 的默认算法是 56 比特的 DES。该加密算法必须被实施，以保证 IPSec 设备间的互操作性。ESP 通过使用消息认证码(MAC)提供认证服务。ESP 既可以单独应用，也可以嵌套的方式使用，还可以和 AH 结合使用。

5.2　信息系统安全策略

5.2.1　信息系统的安全防御策略

信息系统的安全体系属于典型的防御体系，面对各种各样的安全威胁，在构建信息系统防御体系的过程中，应坚持下列原则和策略。

1. 最小特权原则

最小特权原则是信息系统安全的最基本原则，其实质是任何实体只有该实体

需要完成其指定任务的所必需的特权，此外没有更多的特权。例如，如果给一个系统的某个角色指定对某个文件读权限就足以完成需要的任务，那就一定不能给其读写权限。最小特权可以尽量避免将信息系统的资源暴露在网络攻击的安全威胁之下，减少因攻击所造成的破坏。

2. 纵深防御

安全体系既不能仅依靠单一安全机制，也不能进行多种安全服务的简单堆砌，而是应该建立互相支撑的多种安全机制，建立具有一种协议层次和纵向结构层次的完备体系，通过多层机制互相支撑来获取整个信息系统的安全，例如第一道安全闸门防火墙、入侵检测技术、主机安全技术、VPN 技术等。只有这些安全技术构成一个多层次的安全体系，依靠纵深防御策略，才能更好地保证信息系统的安全。

3. 阻塞点

在网络系统对外连接通道上，可以设计监控的连接控制点，系统管理人员在该点对攻击者进行监视和控制。在网络信息安全系统中，位于 Internet 和内网之间就可以构筑防火墙，形成典型的阻塞点，任何对内网的访问操作都必须经过防火墙，以构筑信息安全的第一道安全闸门。

4. 检测和消除最弱连接

系统安全链的强度取决于系统连接最薄弱的环节，木桶理论可以说明这个问题，即一个由木板箍成的桶能装多少水取决于桶壁上最短木板的高度。攻击者只要找出信息安全系统的那个最弱点，并集中力量对其进行攻击，就能突破系统的安全防线。系统安全体系的构建应该注意系统防御中的最弱点，并针对性地采取措施进行加固或者消除。

5. 失效保护

正如在电路系统中，为了提供安全性，当电路出现短路时，空气开关就跳开。在信息系统中的入侵检测系统检测到入侵事件时，必须立即采取应急措施，拒绝入侵者对系统的非法访问，更不允许侵袭者跨入网络内部其他节点，以防止对信息系统造成更大的安全破坏。

6. 防御特色化

通过使用大量不同类型、不同等级的系统来获得额外的安全保护，如果不同的系统安全配置都是一样的，那么入侵者只要知道如何入侵一个系统，对于其他系统的入侵也就轻而易举了。防御特色化就是使用不同厂商的安全产品保护系统，降低因普遍的错误或配置错误而危及系统。

5.2.2 信息系统安全的工程策略

信息系统安全是一个多维、多层次、多因素、多目标的体系，涉及广阔的技术领域。信息系统安全的最终目标是保障信息内容在系统内的任何地方、任何时候和任何状态下的机密性、完整性和可用性。信息系统安全是一个完整的系统概念，其实现必须在安全体系的框架下，采用系统工程的方法进行设计、实施和控制。安全策略应该兼顾信息系统安全的系统性、相关性、动态性和相对性原则。

1. 系统性

只有经过对信息系统进行安全规划，对信息进行优先级保护分类，对信息系统安全脆弱性的分布和强度关系进行分析，对来自内部和外部的威胁手段和技术进行排列，才能准确评估系统的安全风险，建立风险控制模型。这样才能建立起符合自身信息系统实际的并且科学、合理的信息安全体系。

2. 相关性

信息安全系统中各组件之间的关系变化，可能会引起安全风险的强度及分布的变化，因此，安全策略就必须适应这一变化。相关性问题就是要充分考虑并认识到信息系统各组件之间的关系，分析在运行、应用和变更各组件中对安全风险可能产生的相互影响。只有考虑信息安全组件之间的相关性，由此制定的安全策略才是完整的。

3. 动态性

安全策略必须是动态的，即能根据风险的变化对安全策略进行及时的调整，一成不变的静态安全策略在面临信息安全威胁时，会降低其安全作用甚至变得毫无安全作用。因此，安全策略应具备“风险检测—实时响应—策略调整—降低风险”的自适应能力，即信息安全策略应具有动态性。

4. 相对性

信息系统的安全只有相对的安全，没有绝对的安全。理论上说，在计算资源无限的情况下，任何系统都可能被攻破，所以，信息的安全只能是相对的，即使再完善的信息安全方案也有可能面临难以预见的安全问题。因此，安全方案的根本意义不是防范所有违规和网络犯罪，而是在于防范大多数、一般性的违规和常规性的犯罪，同时对恶意违规或犯罪具备探测、记录跟踪、报警和实时反应的能力。这就是信息系统安全的相对性问题。

5.3 信息系统安全框架

本节将从分析 OSI 开放系统互连的安全体系结构着手，讨论信息系统的安全框架。因为基于计算机网络技术的信息系统正是以开放系统互连通信和网络为支撑平

台的，所以，OSI 开放系统互连的安全体系结构可以作为信息系统安全体系的基础。

5.3.1 OSI 开放系统互连安全体系结构

国家标准《信息处理系统开放系统互连基本参考模型——第二部分：安全体系结构》给出了基于 OSI 参考模型的七层协议之上的信息安全体系结构，其核心内容是保证异构计算机进程与进程之间的远距离交换信息的安全，它定义了五大类安全服务，以及提供这些服务的八类安全机制及相应的 OSI 安全管理，并可根据具体系统适当的配置于 OSI 模型的七层协议中。OSI 开放系统安全体系结构如图 5-7 所示。

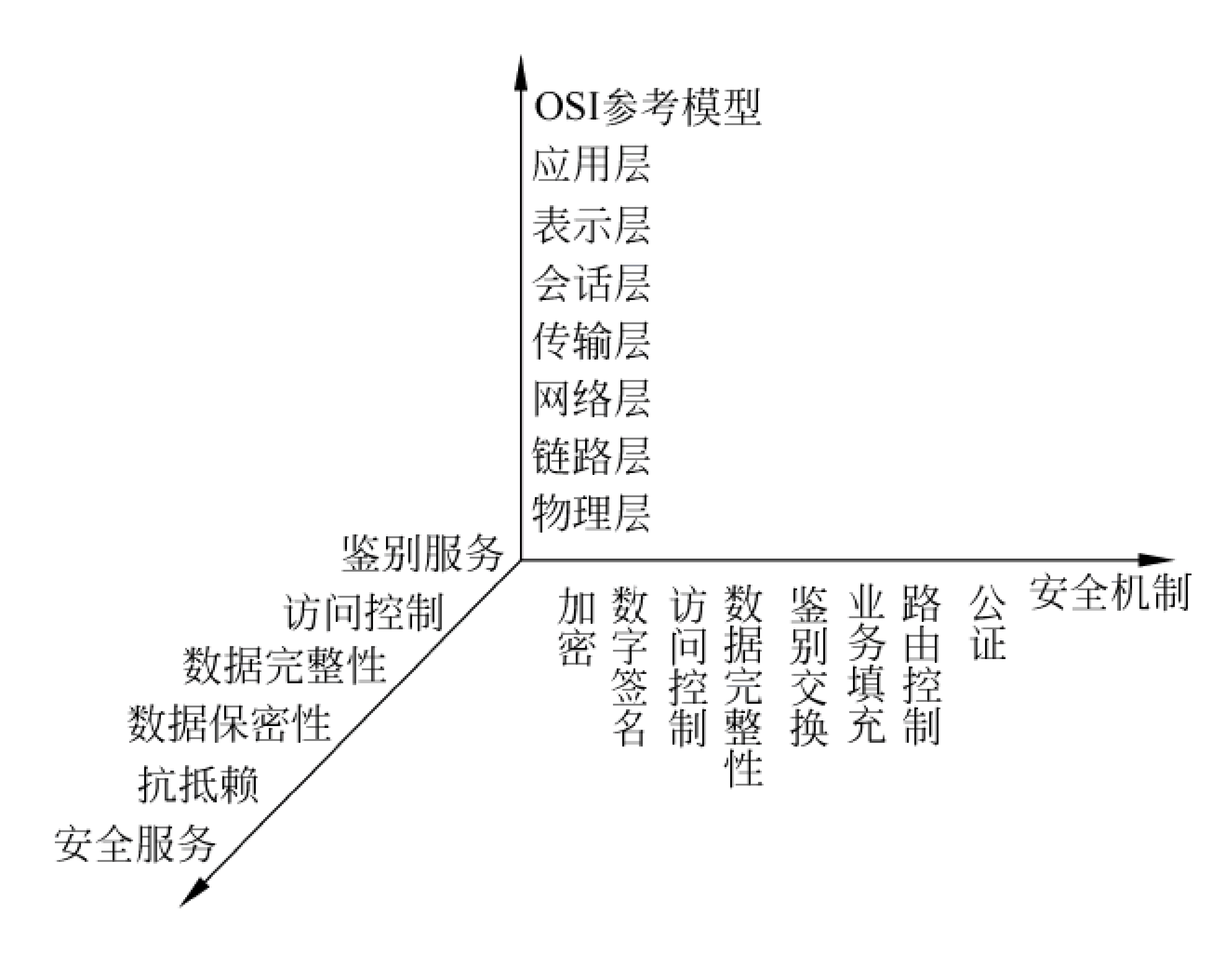


图 5-7 信息安全体系结构

其中，安全体系结构中的一种安全服务可以通过某种单独的安全机制提供，也可以通过多种安全机制联合提供。一种安全机制可用于提供一种或多种安全服务。安全服务可以配置在 OSI 七层协议除了会话层外的每一层上。实际上最适合配置安全服务的是在物理层、网络层、传输层以及应用层上，其他层都不宜配置安全服务。而目前，互联网中的各种安全机制也都是配置在物理层、网络层、传输层以及应用层上。

1. 五类安全服务

五大类安全服务也被称作安全防护措施，即鉴别服务、访问控制服务、数据机密性服务、数据完整性服务、抗抵赖服务。

1）鉴别服务

鉴别服务提供了对通信中对等实体和数据来源的鉴别。其中实体的鉴别意味着，每当某一个实体声称具有一个特定身份的时候，鉴别服务将提供某种方法来证实

这一声明是正确的。鉴别是最基本的安全服务之一,是对付假冒攻击的有效方法。

(1) 对等实体鉴别。

这种服务是在开放系统的两个同层对等实体间建立连接和传输数据期间,为连接实体提供身份鉴别的一种服务。这种鉴别服务可以是单向的,也可以是双向的;鉴别只是鉴别实体身份,并不与实体要进行的访问活动关联起来。对等实体鉴别会产生一个明确的结果,是否允许实体进行通信和其他活动。

(2) 数据源鉴别。

数据源鉴别就是鉴别某个指定的数据项是否来源于某个特定的实体,数据源既不是孤立地鉴别一个实体,也不是为了允许实体执行下一步的操作而鉴别它的身份,而是为了确定被鉴别的实体与一些特定数据项有着不可分割的联系。它对数据单元的来源提供识别,对数据单元的重复或篡改不提供鉴别保护。

2) 访问控制服务

访问控制实现的安全目标是防止对可访问资源进行未授权的访问,包括非授权使用、泄露、修改、破坏和拒绝服务等,其模型如图 5-8 所示。

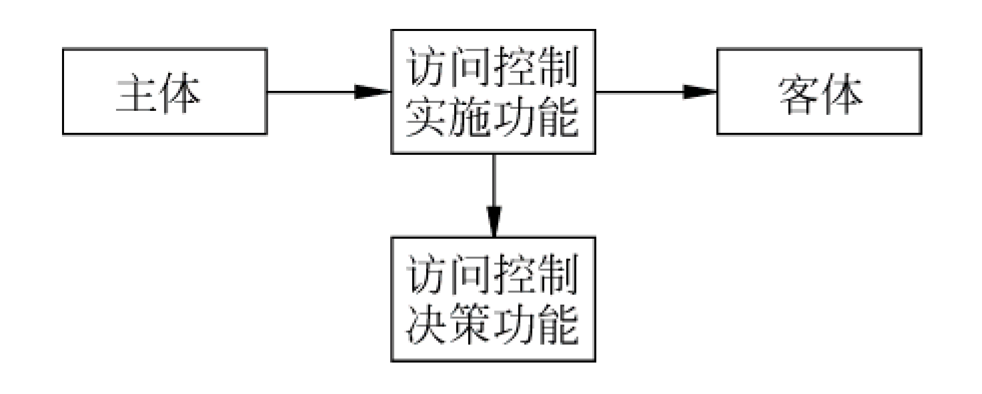


图 5-8　访问控制模型

访问控制策略用于控制实施访问的主体在何种条件下,为了什么目的,可以访问那些客体。在 OSI 访问控制模型中,访问控制实现的安全目标是:

(1) 通过进程对数据、不同进程或其他计算资源的访问控制;

(2) 在一个安全域内的访问控制或跨越一个或多个安全域的访问控制;

(3) 按照其上下文进行的访问控制,根据试图访问的时间、访问者地点或访问路由等因素的访问控制;

(4) 在访问期间对授权更改做出反应的访问控制。

3) 数据机密性服务

机密性服务就是保护信息不泄露给那些未授权掌握这一信息的实体。在信息系统安全中要区分两种类型的机密性服务:数据机密性服务和业务流机密性服务。

数据机密性服务:使得攻击者想要从某个数据项中推出敏感信息是十分困难的。

业务流机密性服务:使得攻击者想要通过观察通信系统的业务流来获得敏感信息是十分困难的。

4）数据完整性服务

完整性服务用于鉴别信息在存储、传输等处理过程中是否受到非授权的修改。

在一次连接中，开始时使用对实体的鉴别服务，并在连接的存活期中使用数据的完整性服务，结合二者就能为数据单元的完整性提供证明。

完整性服务有三种重要类型。

- 连接完整性服务：对连接上传输的所有数据提供完整性保护，以确保接收到的数据和发送的数据一样。
- 无连接完整性服务：对一个无连接数据单元中的所有数据提供完整性保护。
- 选择字段完整性服务：只对某个数据单元中指定的字段提供完整性保护。

此外，完整性服务还可以按照是否具备恢复功能，分为如下两种。

- 具有恢复功能的完整性服务：检测到信息完整性被破坏，并且能正确的将信息恢复到被破坏前的状态；
- 不具有恢复功能的完整性服务：检测到信息完整性被破坏，仅给出报告而不提供信息恢复功能。

5）抗抵赖服务

其他安全服务是针对来自未知者的威胁，而抗抵赖服务的主要目的是保护通信实体免遭来自系统中其他合法实体的威胁，防止通信的任何一方抵赖所进行的传输及传输的内容。

OSI 安全体系结构的抗抵赖服务有两种类型。

（1）有数据原发证明的抗抵赖：为数据的接收者提供数据的原发证明，使发送者不能抵赖这些数据的发送或者否认数据内容；

（2）交付证明的抗抵赖：为数据的发送者提供数据交付证明，使接收者不能抵赖收到过这些数据或否认数据内容。

2. 八种安全机制

OSI 安全体系结构没有说明五类安全服务如何实现，但是它给出了八种基本的安全机制：加密、数字签名、访问控制、数据完整性、鉴别交换、通信业务流填充、路由选择控制和公证机制。可以将一个或多个安全机制配置在适当的（N）层上，用以提供 OSI 安全体系结构的五类安全服务。

1）加密机制

加密机制是各种安全服务和其他安全机制的基础，既能为数据提供机密性，也能为通信业务流信息提供机密性，并且还能成为其他安全机制中的一部分。OSI 安全体系结构中的加密机制包括三方面的内容，包括加密层的选取、加密算法的类别、密钥的管理。

2）数字签名机制

数字签名是在数据单元上附加数据，或对数据单元进行的密码变换，使得接收者能够证实数据单元的来源及其完整性，实现对数据的保护。

数字签名机制需要确定两个过程：对数据单元签名和对签过名的数据单元进行验证。对数据单元的签名，即使用签名者所私有的（独有的和机密的）信息作为私钥，对数据单元进行加密，或产生该数据单元的一个密钥校验值。

对签过名的数据单元进行验证，即使用公钥证书信息来验证该签名是不是由用签名者的私有信息产生的。

数字签名机制的本质特征为该签名只能使用签名者的私钥才能生成。因此，当该签名得到验证后，它能在事后的任何时候向第三者证明——只有那个私钥的唯一拥有者才能产生这个签名。

3）访问控制机制

访问控制机制是被用来实施对资源访问或操作加以限制的策略，把对资源的访问只限于那些被授权了的用户，访问控制可以应用于通信联系中的任一端点或者中间点。

为了确定一个实体的访问权并实施该访问权，访问控制机制可以使用该实体的已鉴别的身份，或使用该实体的有关信息，或应用该实体的权力。如果这个实体试图访问非授权资源，或者非法使用授权资源，那么访问控制机制可以阻止这一企图的实施。另外还可以产生报警信号或记录，将它作为安全审计跟踪的一个事件。

4）完整性机制

完整性机制的目的是保护数据，以避免未授权的数据乱序、丢失、重放、插入和篡改。提供完整性机制的最基本手段是使用哈希函数提取要发送的消息的摘要信息，将摘要信息作为发送消息的一部分。如果消息由于各种原因导致乱序、丢失、重放、插入或篡改，消息的接收方都可以通过再次计算消息的摘要值和收到的摘要值进行比较而检验消息的完整性。

5）认证交换机制

认证交换技术的选用取决于使用它们的环境，在许多场合，它们必须与下列各项技术结合使用：

（1）时间标记与同步时钟；

（2）两次握手（单向认证）和三次握手（双向认证）；

（3）数字签名和公证，用于抗抵赖服务。

6）通信业务填充机制

通信业务填充机制也是提供机密性的一个基本机制，它属于一种反分析的技术。它包括生成伪造的通信实例、伪造的数据单元、伪造数据单元中的内容，以此

将协议数据单元填充到一个固定的长度，以防止对通信业务的分析，对通信机密性提供保护。

7）路由选择控制机制

路由选择控制机制使得路由能动态的或预定的选取，以便通过物理上安全的子网络、中继站或链路来通信，确保敏感数据仅在具备适当保护级别的路由上进行传输。

8）公证机制

公证机制是由可信的第三方公证人提供数据完整性、数据源、时间和目的地等的公证和保证。当采用公证机制时，数据便在参与通信的实体之间经由受保护的通信实体和公证方进行通信。

OSI 安全体系中安全服务与安全机制之间的关系如表 5-1 所示，其中每一种安全服务可以由一种或多种安全机制联合提供。

表 5-1　安全服务和安全机制之间的对应关系

安全服务		安全机制							
		加密	数字签名	访问控制	完整性	鉴别交换	业务填充	路由控制	公证
鉴别	对等实体鉴别	√	√			√			
	数据源鉴别	√	√						
访问控制				√					
机密性	连接机密性	√						√	
	无连接机密性	√						√	
	选择字段机密性	√							
	通信业务流机密性	√					√	√	
完整性	带恢复的连接完整性	√							
	不带恢复的连接完整性	√			√				
	选择字段连接完整性	√			√				
	无连接完整性	√	√		√				
	选择字段无连接完整性	√	√		√				
抗抵赖	有数据原发证明的抗抵赖		√		√				√
	交付证明的抗抵赖		√		√				√

5.3.2　TCP/IP 安全体系

1. TCP/IP 安全体系结构

前面介绍的 OSI 开放互连体系是从现实应用的各种网络中提取出较为抽象的一般共性而形成的国际标准。下面的 TCP/IP 安全体系结构则以 TCP/IP 协议为技术支撑，TCP/IP 协议是目前业界公认的事实上的工业标准。基于 TCP/IP 协议

的网络体系结构如图 5-9 所示。

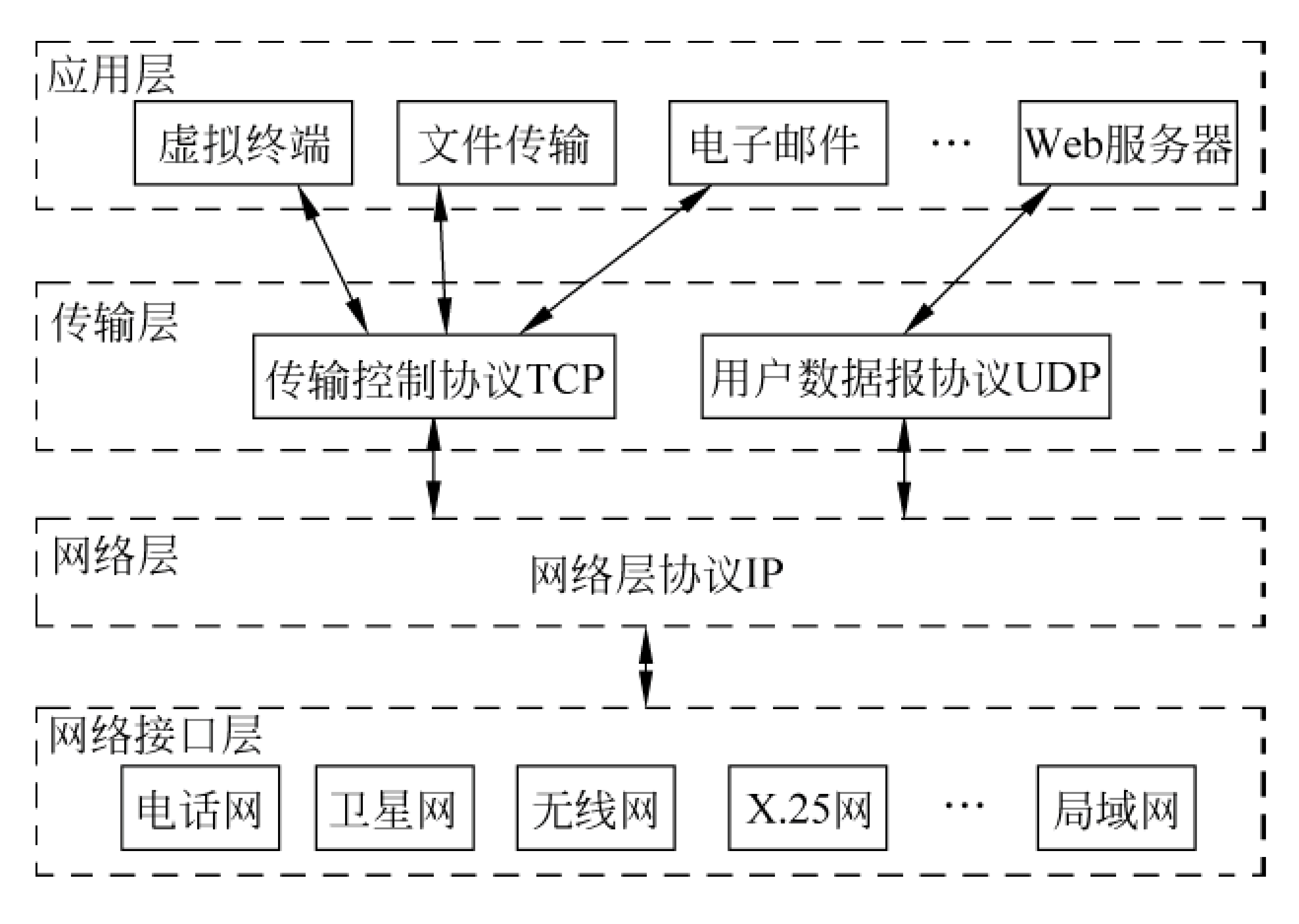


图 5-9　基于 TCP/IP 协议的网络体系结构

作为一种实际上广泛应用的协议集，TCP/IP 也完全可以用 OSI 参考模型来解释。由于 OSI 参考模型与 TCP/IP 协议集模型之间存在这种对应关系，因而可以根据 OSI 安全体系结构框架，将各种安全机制和安全服务映射到 TCP/IP 的协议集中，形成一个基于 TCP/IP 协议层的网络安全体系结构，如表 5-2 所示。

表 5-2　基于 TCP/IP 协议层的网络安全体系结构

安全服务		TCP/IP 协议层			
		网络接口层	网络层	传输层	应用层
鉴别	对等实体鉴别		√	√	√
	数据源鉴别		√	√	√
访问控制			√	√	√
机密性	连接机密性	√	√	√	√
	无连接机密性	√	√	√	√
	选择字段机密性				√
	通信业务流机密性	√	√		√
完整性	带恢复的连接完整性			√	√
	不带恢复的连接完整性		√	√	√
	选择字段连接完整性				√
	无连接完整性		√	√	√
	选择字段无连接完整性				√
抗抵赖	有数据原发证明的抗抵赖				√
	交付证明的抗抵赖				√

2. 网络层安全协议 IPSec

IPSec 是 IP Security 的缩写，即 IP 层安全，它基于两种关键技术：加密封装和报文认证。IPSec 是一套协议包，它由两大部分组成：

(1) 密钥交换协议 IKE，用于建立安全的分组流；

(2) 保护分组流的协议，包括封装安全载荷协议(ESP 协议)或认证报头协议(AH 协议)，用于保证分组流的机密性、来源可靠性(认证)、无连接的完整性，并提供抗重播服务。

IPSec 把多种安全技术集合到一起，可以建立一个安全和可靠的隧道，提供 IP 层的安全服务，可在主机或网关上实现。IPSec 数据包结构如图 5-10 所示，是在 IP 包头后增加几个新的字段来实现安全保证。

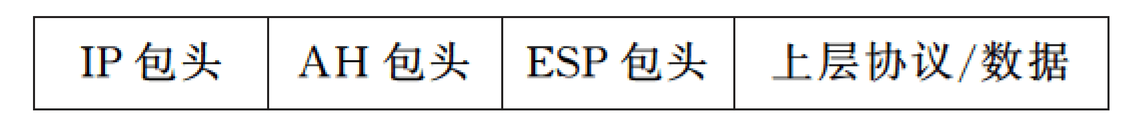


图 5-10 IPSec 数据包结构

IPSec 包括三个基本协议：

(1) 认证报头协议(Authentication Header，AH)——为 IP 报头提供信息源验证和完整性保证。

(2) 封装安全载荷协议(Encapsulating Security Payload，ESP)——提供加密保证。

(3) 密钥交换协议(Internet Key Exchange，IKE)——提供双方交流时的共享安全信息。

IPSec 各部件之间的关系如图 5-11 所示。

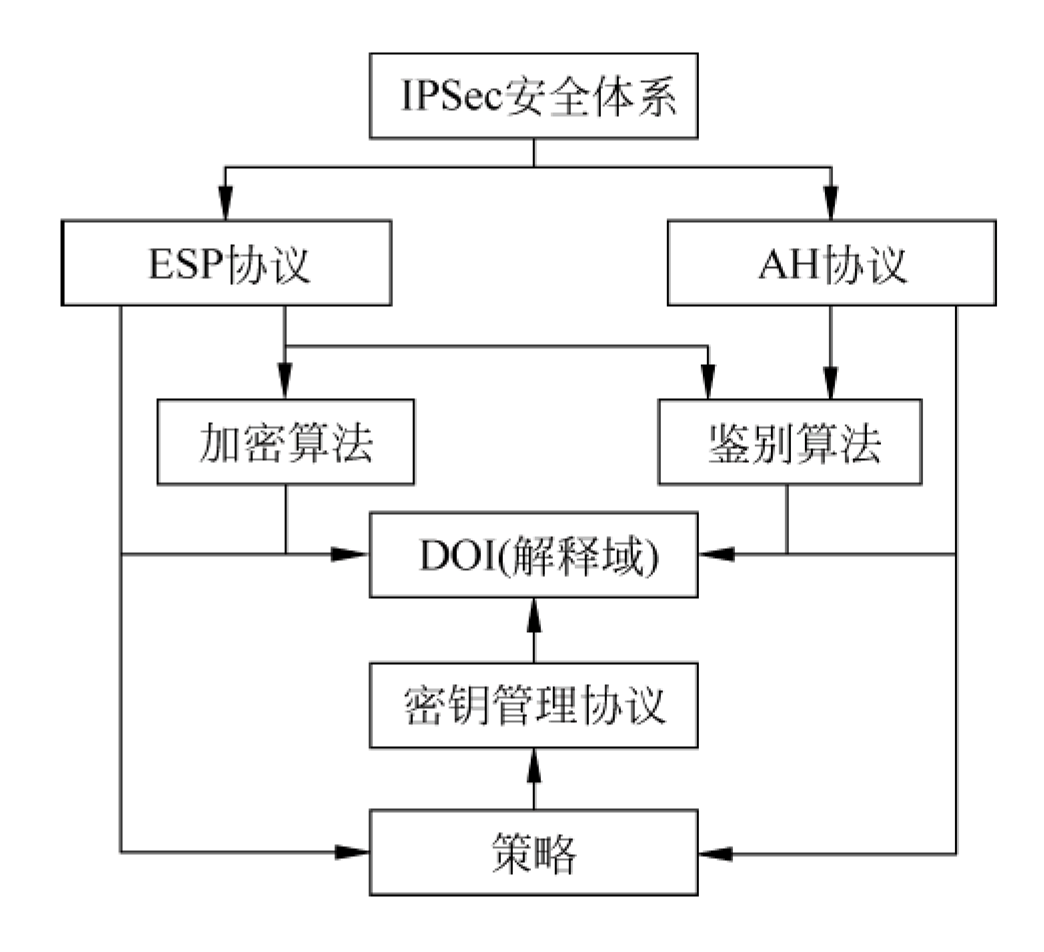


图 5-11 IPSec 各部件之间的关系

ESP 协议规定了为通信提供机密性和完整性保护的具体实现方案，为 IP 数据包提供数据源验证、数据完整性、抗重放和机密性服务，涉及加密和鉴别算法。

AH 协议为 IP 数据包提供数据完整性和验证服务，涉及鉴别算法。

为了 IPSec 通信两端能互相交互，通信双方必须保持对通信消息相同的解释规则，即拥有相同的解释域(Domain of Interpretation，DOI)。IPsec 定义了两种安全机制 AH 和 ESP，并以 IP 扩展头的方式增加到 IP 包中。

5.3.3　OSI 安全体系框架

信息系统安全问题仅仅依靠技术手段解决是不充足的，要想有效地保护系统的安全，最大限度地减小所面临的安全风险，还必须从安全技术、组织机构与人事管理等方面采取综合措施。OSI 安全体系框架就是在 OSI 安全体系结构基础上构建的，主要包括技术体系、组织体系和管理体系三种安全体系。OSI 安全体系框架如图 5-12 所示。

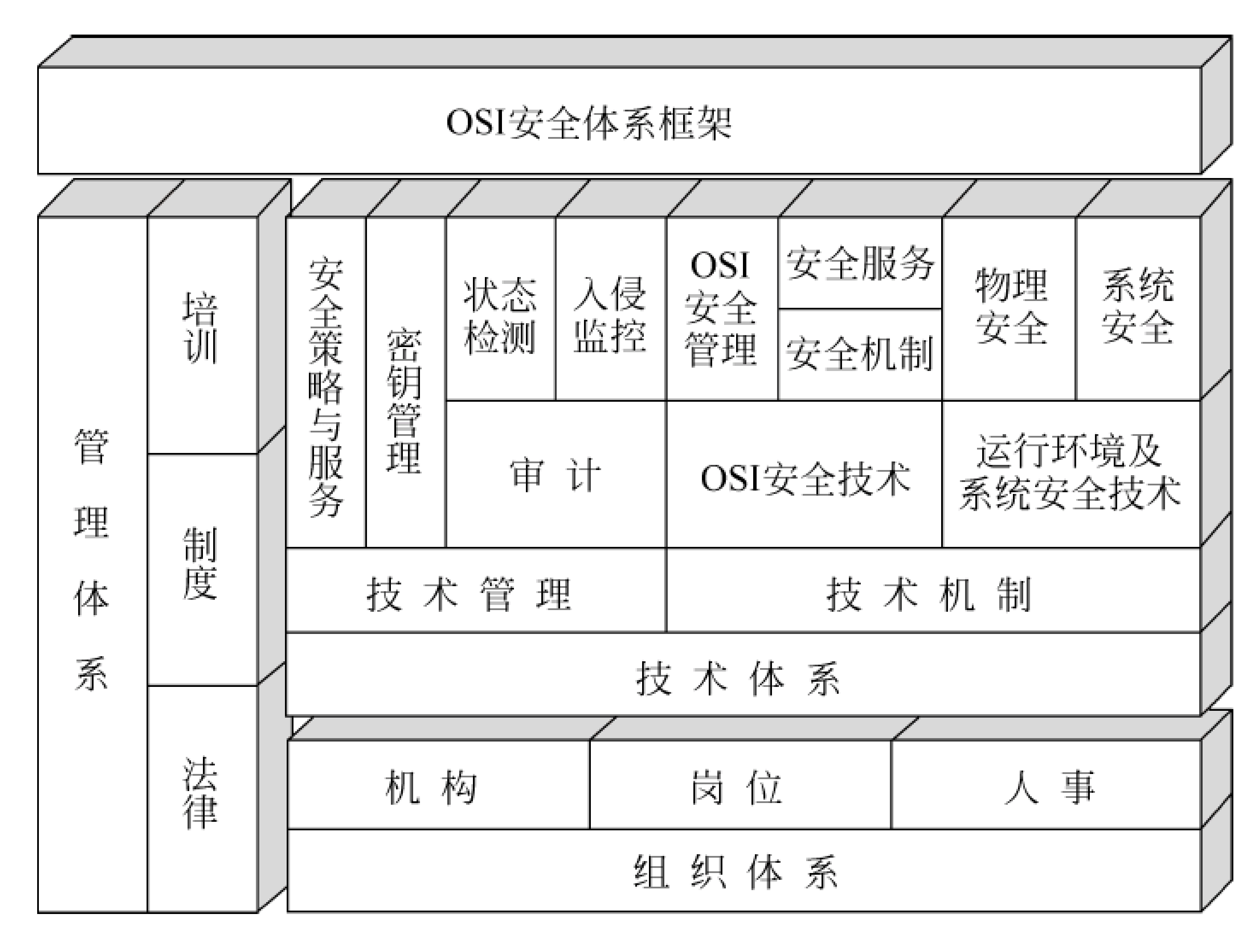


图 5-12　OSI 安全体系框架

1. 技术体系

技术体系是全面提供信息系统安全的技术保障系统。OSI 安全体系结构通过技术管理将技术机制提供的安全服务，分别或同时应用在 OSI 协议层的一层或多层上，为数据、信息内容、通信连接提供机密性、完整性和可用性保护，为通信实体、通信连接、通信进程提供身份鉴别、访问控制、审计和抗抵赖保护，这些安全服务分别作用在通信平台、网络平台和应用平台上。通过对信息系统与安全相关组件的操作系统的安全性选择措施或自主控制，使信息系统安全组件达到相应的安全

等级。

OSI 安全体系中的技术体系框架设计，可以借鉴美国的 DISSP 计划提出的三维安全体系思路，将协议层、系统构成单元和安全服务（安全机制）分别作为三维坐标体系的三维，如图 5-13 所示。

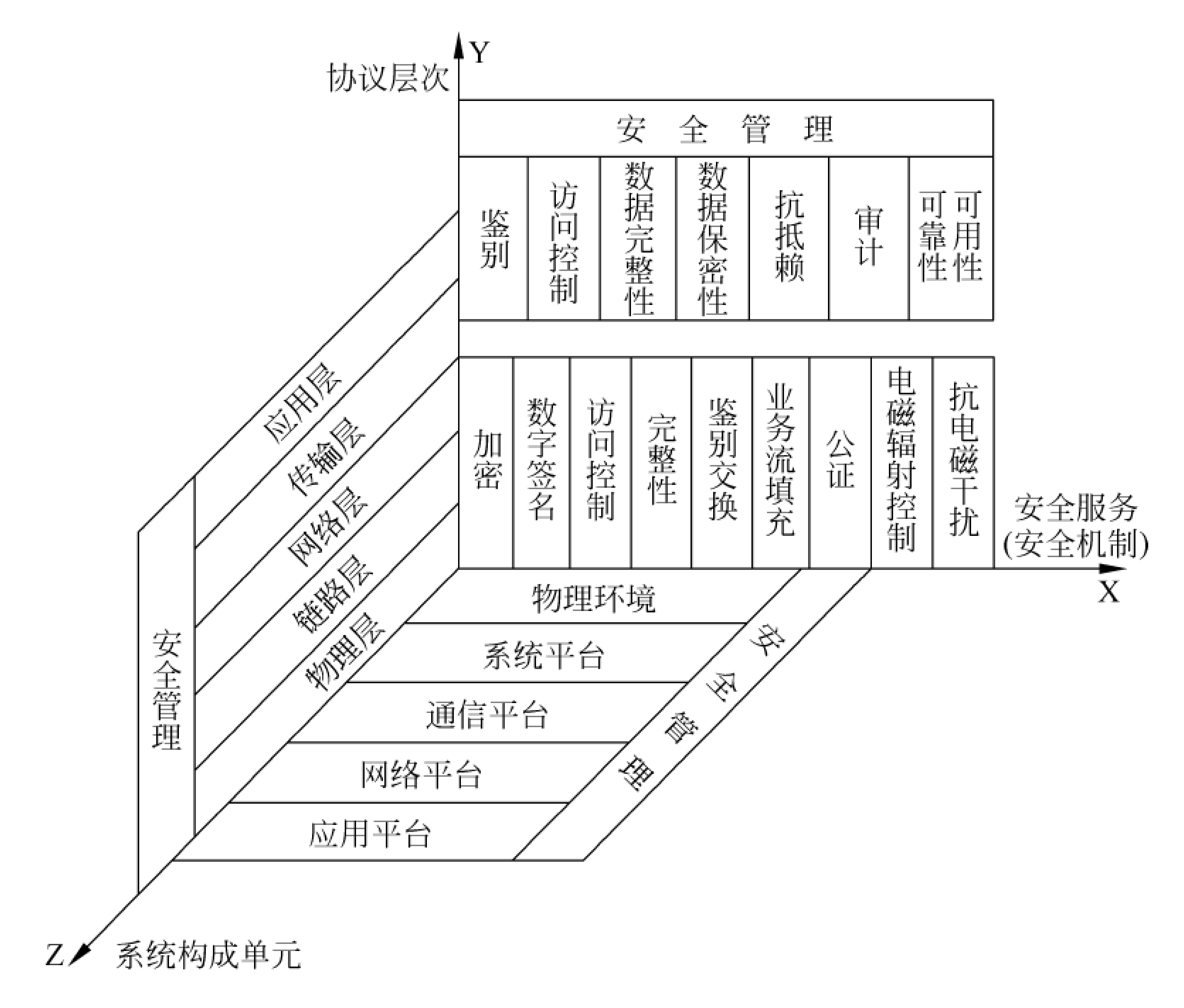


图 5-13　安全技术体系结构

安全服务（安全机制）作为 X 维，协议层作为 Y 维，系统的构成单元作为 Z 维。在 X 维中，安全机制既不直接配置在协议层，也不直接作用于系统单元，而是通过提供安全服务发挥作用。Y 维的协议层只选取可适当配置安全服务的五个层次。

2. 组织体系

组织体系是信息系统安全的组织保障系统，由机构、岗位和人事三个模块构成一个体系。

机构的设置包括决策层、管理层和执行层三个层次。决策层作为信息系统的主体，是决定信息系统安全的重大事宜的领导机构，由主管信息工作的负责人为主要责任人，并有国家安全、公共安全、机要和保密职能部门的负责人及信息系统主要负责人共同参与组成。管理层是日常管理机关，根据决策机构的决定进行全面规划并协调各个方面的力量确定信息系统安全的实施方案，制定、修改安全策略，处理安全事故等。执行层是在管理层协调下，具体负责特定安全事务的逻辑群体，这个群体可以分布在信息系统的各个岗位上。

岗位是根据系统的安全需要而设定的负责安全事务的职位，岗位在系统内部可以是具有垂直领导关系的若干层次构成的一个序列，一个人可以负责一个或几个安全岗位，但是一个人不能同时兼任安全业务岗位及该岗位所对应的系统管理工作。

3. 管理体系

一般说来，信息安全目标的达成“三分靠技术、七分靠管理”，由此可见，信息安全管理的重要性，可以说管理是信息系统安全的灵魂。信息系统安全的管理体系可以划分为法律管理、培训管理及制度管理三部分。

法律管理是指根据国家相关法律和法规对信息系统主体及其行为进行规范和约束。法律管理具有对信息系统主体的行为的强制性约束，并且具有明确的管理层次性。安全方面的法律和法规是信息系统安全工作应该遵守的最高行为准则。

制度管理是信息系统内部制定的一系列内部规章制度，主要包括安全管理和执行机构的行为规范，岗位设定及其操作规范，岗位人员的素质要求及行为规范，以及内部关系与外部关系的行为规范等。制度管理是法律管理的形式化、具体化，也是法律、法规与管理对象之间的接口。

培训管理是确保信息系统安全的前提，其内容包括法律法规培训、内部制度培训、岗位操作培训以及与岗位相关的重点安全意识相结合的培训、业务素质与技能技巧培训等，培训的对象几乎包括信息系统有关的所有人员（不仅仅是从事安全管理和业务的人员）。

5.3.4　信息系统安全体系框架

根据信息系统安全的概念，信息系统安全的总需求包括物理安全、网络安全、数据安全、信息内容安全、信息基础设备安全与公共信息安全等诸多方面；安全的最终目标是保证信息的机密性、完整性、可用性、可审计性和抗抵赖性，以及信息系统主体对信息资源的控制。从这里给出的总需求上看，其中网络安全、数据安全、信息内容安全等可以通过 OSI 安全体系提供的安全服务、安全机制及其管理获得。通过分析 OSI 安全体系框架，系统性的完整的构建信息系统的安全体系框架如图 5-14 所示。

信息系统安全技术体系是实现安全信息系统所采用的安全技术的构建框架，包括信息系统安全的基本属性、信息系统安全的组成与相互关系、信息系统安全等级划分、信息系统安全保障的基本框架、信息系统风险控制及其技术支持等。如图 5-15 所示为信息系统安全技术体系框架，这一框架是在 OSI 安全体系框架的技术体系基础上构建的。

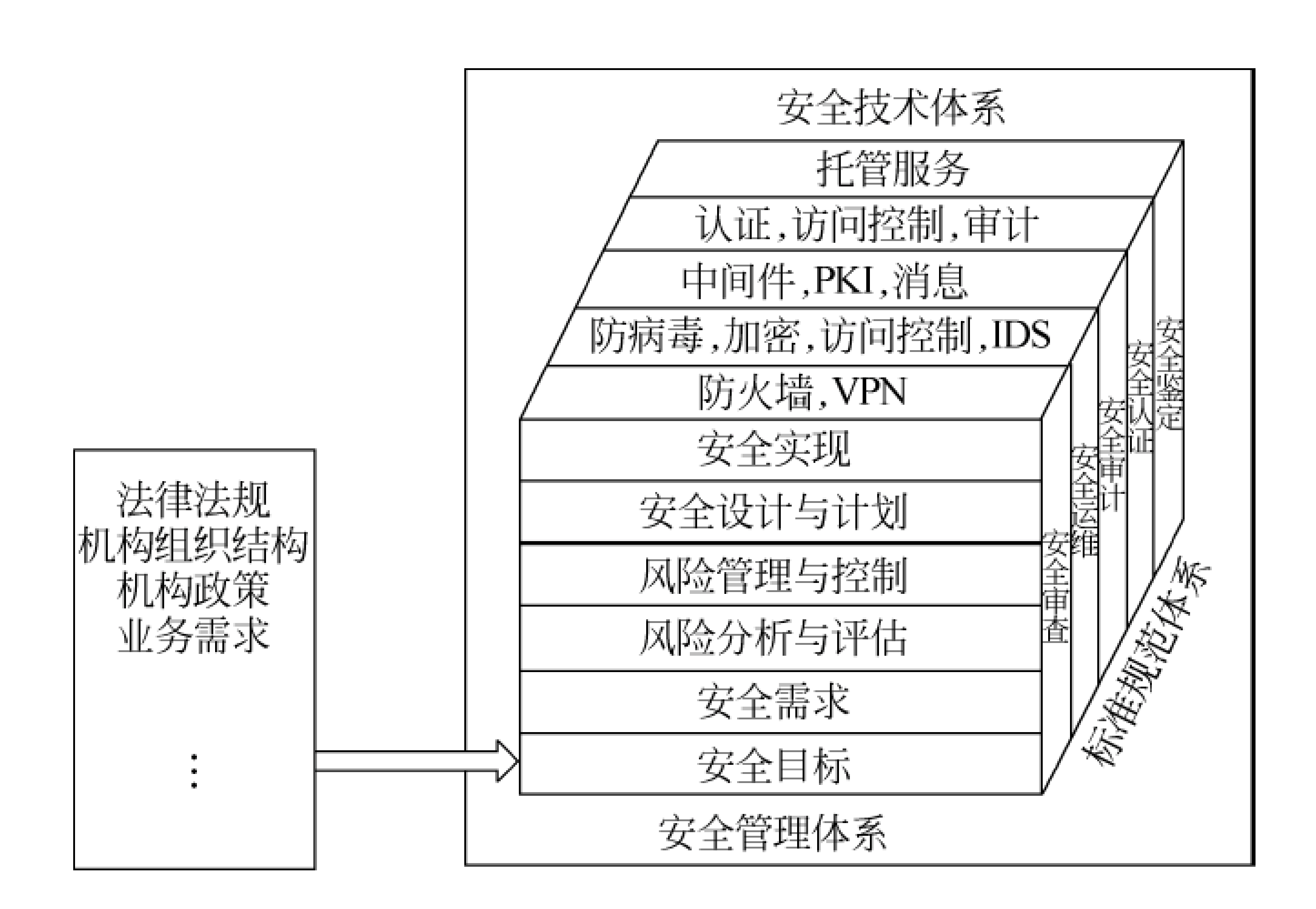


图 5-14　信息系统的安全体系框架

<table>
<tr><td colspan="4">安全信息系统</td></tr>
<tr><td>应用安全</td><td rowspan="5">防火墙/
VPN</td><td rowspan="5">入侵检测</td><td rowspan="5">身份识别/
权限管理</td></tr>
<tr><td>传输安全</td></tr>
<tr><td>访问控制</td></tr>
<tr><td>安全协议</td></tr>
<tr><td>加密算法</td></tr>
<tr><td colspan="4">操作系统安全</td></tr>
<tr><td colspan="4">硬件安全</td></tr>
</table>

图 5-15　信息系统安全技术体系框架

5.4　主机安全技术

随着网络的日益普及,信息安全事件不断涌现,安全威胁的阴霾日渐浓厚。而主机作为数据存储、系统运行的实体环境,在“攻击之矛”与“防护之盾”之间的较量中,扮演着越来越重要的角色,主机安全逐渐成为新型网络战中的战略高地。攻击者采用的入侵方式可能是多种多样的,比如传播病毒、利用漏洞植入“后门”、暴力破解等等,但最终目标通常只有一个,就是主机以及主机上的数据和核心业务系统。对用户来说,主机也是信息网络中的核心设备,只有保护好主机,才能真正实现信息安全。

5.4.1　主机安全威胁分析

当前人们关注到的很多安全威胁都是从网络上发起的，网络安全事件也确实占到安全事件的大部分。但是应该注意的是，网络攻击的目标不仅仅满足于获取个人信息，而是商业机密、国家秘密等所有对攻击者有利的数据，防守网络攻击的最后一道防线就在主机系统。服务器作为信息系统中重要的主机计算节点，是信息系统的核心。无论是在C/S计算模式还是B/S计算模式中，服务器都是应用服务的运行中心和数据处理中心，也是信息系统中敏感信息的直接载体，因此对服务器的保护是保证整个信息系统安全的基础。目前用户对基于网络应用的外部防范关注较多，却往往忽略了服务器自身安全，然而在信息系统中最薄弱、最易受攻击而保护措施又最缺乏的就是对服务器的保护。

由于所有的数据最终都是要通过主机系统来进行处理、存储，这个环节一旦受到攻击，整个信息系统中最有价值的部分就成为入侵者可以直接获取的有价值的资源了。主机系统一旦宕掉或者出现安全问题，对于整个信息系统造成的损失可能是巨大的。下面列举一些常见的对主机的攻击方式。

1. 端口扫描

端口扫描是黑客们用来决定主机的哪一个端口可以被他们利用的流行手段，这个动作往往是黑客入侵的第一步，也是主机安全系统作用的第一个环节。端口扫描是攻击者向每个端口发送一个消息的过程，根据端口的回应来确认该端口是否可用和更进一步侦察它的漏洞。

2. MAC(Media Access Control)地址欺骗

一些黑客会使用MAC地址欺骗来窃取一段网络对话，以达到控制其中一台主机的目的。MAC地址用来唯一标识一台主机、服务器、路由器以及其他设备。当两台主机之间需要通信的时候会发送一个ARP(Address Resolution Protocol)包，MAC地址欺骗就是攻击者利用伪造ARP包的方式窃取通信另一端的数据。

3. IP 地址欺骗

IP地址欺骗与MAC地址欺骗类似，也成为黑客常用的一种窃取计算机之间对话的手段。例如，计算机A正在与计算机B进行通信时，黑客先向计算机A发送一个数据包，使其退出对话，然后假装计算机A跟计算机B进行通信，达到窃取对话内容，进而攻击计算机B的目的。

4. DoS 攻击

拒绝服务攻击(Denial of Service，DoS)是目前常见而又难于防范的一种攻击手段，对于服务器尤其明显，DoS攻击的目的是阻止合法的用户使用网络服务，造

成网络拥塞，使得某些网络服务变得不可用。通常表现在如下几个方面：

（1）制造大流量的无用数据，使得通往被攻击主机的网络拥塞，造成被攻击的主机无法与外界正常通信。

（2）利用被攻击主机提供的服务或者传输协议上的处理重复连接的缺陷，反复高频发出重复服务请求，使被攻击主机不能及时的处理其他的正常请求。

（3）利用被攻击主机的服务程序或传输协议本身的实现缺陷，重复发送畸形的攻击数据，使得系统错误地分配大量系统资源，直至主机处于挂起状态甚至造成被攻击主机死机。

5. DLL 和 Windows 钩子

这是恶意程序经常利用的一种主机攻击的手段。Windows 系统是建立在事件驱动的机制上的，即整个系统都是通过消息的传递来实现的。而钩子（Hooks）是 Windows 系统中重要的系统接口，利用它可以截获并处理送给其他应用程序的消息，完成普通应用程序难以实现的功能。钩子机制可以用来监视系统或进程中的各种事件消息，截获发往目标窗口的消息并进行处理，利用这种机制，通过在系统中安装自定义的钩子，就可以实现监视系统中特定事件的发生，完成特定的功能，例如截获键盘、鼠标的输入，屏幕取词以及日志监视等。此外，在创建 Windows 程序时，链接过程并不把 DLL（Dynamic Link Library 动态链接库）文件链接到程序上，而是在程序运行并调用一个 DLL 中的函数时，该程序才要求这个函数的地址。此时 Windows 才在 DLL 中寻找被调用函数，并把它的地址传送给调用程序。所以通过在 DLL 中插入自己的函数入口，从而实现对于特定事件的监控。

以上针对主机的安全威胁都是利用了操作系统和网络协议的一些可利用的漏洞，对主机的正常工作进行干扰和破坏。

5.4.2 主机安全防护技术

根据国际标准化委员会对信息安全给出的定义，服务器安全应归结到三个层面，即物理层面（服务器硬件）、运行层面（操作系统与软件）及数据层面。

物理层面的安全，主要是指主机系统及其附属设备的安全，是大型计算机安全的基础。计算机机房地理环境的选择，各种设施的安全位置，防火、防水、防震、防雷、防静电、防电磁干扰、防盗、防尘等方面都有一整套比较成熟的规范和参数，大型计算机在物理层面比一般计算机要求更加严格。

运行层面的安全，主要是指保证大型主机软件的完整性，堵塞软件中存在的漏洞，防止软件的非法复制、泄露和修改；而操作系统安全是保障主机安全的基础，或者说安全的操作系统是整个信息系统安全的基础。如果信息安全框架的构造只关注网络层面的防护，而忽略操作系统内核安全这一问题，就如同将坚固的堡垒建

立于沙丘之上,安全隐患是巨大的。从主机节点的操作系统实施安全防范,可以将不安全因素从源头进行控制。因此,对主机节点进行安全加固处理是十分必要和重要的。

数据安全,主要是指保证存储在大型计算机系统中或在大型主机系统间或大型主机与其他计算机系统间传输的数据不受非法删改或意外事件的破坏,保证敏感信息的机密性,使得敏感数据信息不被恶意攻击者窃取。数据安全是主机安全的核心,实现数据安全的基础是各种信息加密机制。

以保障终端(特别是服务器)操作系统安全为基础,以服务器运行安全、数据安全、安全管理三个方面为目标,实现服务器全生命周期的安全防护。这是立足计算节点实现计算环境安全的根本目标,同时也是从信息安全的根本源头构建主动防御体系的重要手段。

我国在信息技术方面起步较晚,核心技术方面一直受制于西方国家,重要信息系统和基础信息网络大量使用国外基础软件以及核心关键设备,这种局面使得我国信息安全遭受到各种未知的后门、漏洞的威胁,严重影响到国家安全。

保障主机安全通过两条途径:一是自主可控,二是安全可靠。自主可控是前提,安全可靠是保障。自主可控是指主机系统的国产化,从硬件逻辑、操作软件源代码及系统运维层面实现自主可控,消除国外主机系统的安全威胁。事实上,在关键信息系统中强制性使用本国产品,已经是美国、欧洲、日本等发达国家和地区的惯例。

为什么一定要自主可控?信息安全专家认为,美国很多关键信息基础设施在私营大企业手里运营,这些企业和美国政府之间的关系很好,长期配合美国政府提供一些服务,2013年的斯诺登事件所披露的信息就表明了这一点。而这些大企业中,有些已在中国经营了很长时间,其产品遍布我国的政府机关、重要企业的各个部门。

有国外媒体报道,美国国安局曾与美国加密技术公司RSA达成1000万美元的协议,要求后者在移动终端普遍使用的加密技术中放置后门,以便让美国国安局通过后门程序轻易破解各种加密数据,获取信息。虽然RSA很快对此予以了否认,但是该消息仍然让人们怀疑有企业协助美国政府对他国网络空间进行监控。

目前,我国的浪潮等公司启动了主机安全战略,并发布集硬件、操作系统、安全软件"三位一体"的主机安全方案,其核心在于依托自主创新,发展安全可控的主机安全关键技术及系列产品、方案。针对主机安全领域,将在高速互联芯片/固件的安全设计技术、主板的安全设计技术、虚拟化安全技术、可信计算应用技术和主机安全性评估技术等方面进行自主研发,实现全线产品的安全化。

5.5 云计算安全

随着云计算技术的迅速发展和应用,安全问题的重要性呈现逐步上升的趋势,成为制约其发展的重要因素。在云计算环境下,数据被集中存储在云端,由云数据中心的管理者对存储的信息资源进行统一管理、统一分配,并实现均衡负载、软件部署以及安全控制,并对云计算系统进行可靠的安全实时监测,从而使用户的数据得到最大限度的安全保证。然而,集中管理的云计算中心必将成为黑客攻击的重点目标,同时,对于客户来说,云计算服务商也是数据安全的防范对象之一。由于云计算环境的巨大规模及其前所未有的开放性和复杂性,云计算系统面临着比传统的主机与网络环境更为严峻的信息安全考验。数据的私密性、安全性,以及云计算服务的稳定性已成为用户是否使用云服务的关键性指标。

5.5.1 云计算的基本特征与安全问题

云计算具有五个最基本的特征:一是按需自助服务,即用户可以根据需求对计算资源进行单边部署而自动获取,无须云服务提供商进行人工配合;二是网络连接泛在化,即云计算资源可以通过无处不在的网络方便地获取,访问机制能够使用户方便地通过异构的客户平台使用云计算服务;三是资源池的地理无关性,是指云计算服务商采用多租户模式,根据用户需求对物理资源和虚拟资源进行动态地分配和再分配,而用户不必知道资源的所在的位置;四是资源部署的快速灵活,即云计算供应商可快速地扩充和缩小资源,对于用户而言,云计算资源通常可以认为是无限的,即可以在需要的时间购买任何数量的资源;五是服务计费,即云计算系统能自动控制和优化服务资源,并按照不同类型的服务进行计费。

在以上基本特征中,资源的虚拟化和服务化可以说是云计算的最重要特征。在云计算模式下,信息技术所需的基础设施不再是用户必须在本地配置的,而可以通过租用云服务商所提供的服务来满足应用需求。在云计算框架中,各层次的功能被封装成抽象实体,通过虚拟化技术实现这些服务,并对用户提供各层次的云服务。虚拟化技术将底层的硬件,包括服务器、存储与网络设备等全面虚拟化,基于虚拟化技术,通过建立一个可以随需而选的资源共享、分配、管控平台,根据上层的数据及业务形态的不同需求搭配出多种相互独立的应用,形成一个可伸缩的、服务为导向的 IT 基础架构。

云计算的以上特征可能带来诸多新的安全问题,问题的根本原因在于用户对数据和环境不再拥有完全的控制权。由于云计算打破了地域的概念,数据不再是存放在某个确定的物理节点上,而是存放在由云服务商动态地提供的存储空间中,

这些存储空间可能是现实的，也有可能是虚拟的，可能分布于不同的国家和区域；由于用户对存储于云中的数据不具有完全的管理权，相对于传统的本地数据存储及处理方式，云服务器中的数据存储和处理变得更加不可控；云计算环境下，各类云应用之间不存在固定不变的基础设施及安全边界，数据的安全不是依靠机器或网络的物理边界来保障，而是由云服务提供商负责，因此，在云计算环境中，用户数据安全与隐私保护成为重要的研究课题。

另一方面，云计算中大量采用虚拟化技术，虚拟平台的安全也是云计算安全的重要方面。由于功能与性能的不断提升，虚拟化平台变得越来越庞大，也越来越复杂，安全管理的难度也在随之增大。当前，虚拟平台的安全漏洞不断地涌现，如果这些漏洞被黑客利用而获得虚拟平台的管理软件控制权，云计算系统的安全将会受到直接威胁。

再者，云计算中的多层服务模式也将引发安全问题。服务专业化是云计算发展的趋势之一，即一个云服务商在对外提供各种服务的同时，也许需要购买一些其他云服务商所提供的服务，因而一个用户所享用的云服务可能间接地涉及多个云服务提供商，这种服务的多层转包无疑极大地提高了问题的复杂性，也增加了安全方面的风险。

以上分析表明，在云计算情况下，传统的安全域划分、网络边界防护等安全机制已难以满足系统的安全需求。用户数据的安全、用户隐私信息的保护、数据异地存储服务的安全等诸多安全问题都需要关注。概括起来说，云计算安全涉及数据安全、应用安全、虚拟化安全三方面的内容。

根据云计算平台的特点，基于通用安全措施和云计算的服务模型，云计算安全总体框架可以构建如图 5-16 所示。

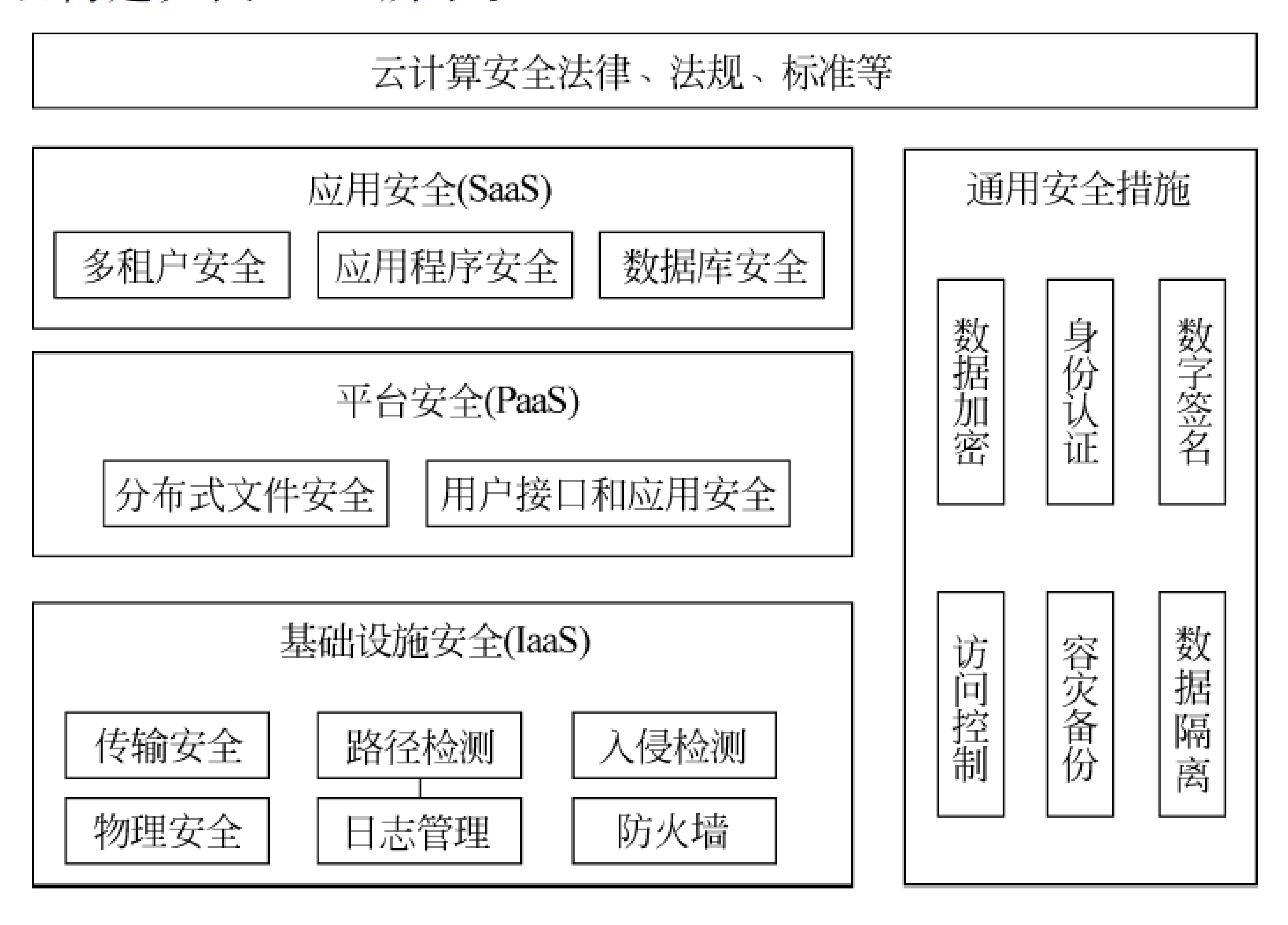


图 5-16　云计算安全总体框架

传统安全技术发展了多年，相应的标准、法律、法规也都比较成熟，但现在的云计算安全缺少标准，其中的政策、法规也不健全；再加上云计算自身的特点，数据可以存储在世界上任何一个国家，当出现问题时，不同国家的政策不同也是云计算安全需要面对的一个挑战。目前，关于云计算安全标准的一些研究也都处于进行中，尚未获得一致认可的安全技术和标准。

根据云计算技术体系架构，云计算的安全从高到低分为应用安全、平台安全和基础设施安全三个层次的安全问题。

5.5.2 IaaS 层安全问题及措施

IaaS 层处于云计算平台的最底层，负责为上层的云应用提供安全的数据存储和安全的计算等 IT 资源服务，IaaS 层的安全是整个云计算安全体系的基石。IaaS 平台的安全问题既有传统的数据中心安全特性，也存在自身特有的安全风险。

首先，与传统的数据中心一样，IaaS 平台面临着各种传统的安全问题。例如，在物理层考虑厂房安全等物理安全；在存储层需要考虑数据加密、备份、归档、灾难恢复等问题；在网络层的 DDoS 攻击、数据传输机密性等风险；在数据层则存在数据库安全、数据的隐私性与访问控制问题等；在应用层则需要考虑程序完整性检验、访问控制与漏洞管理等问题。

其次，由于 IaaS 平台上大量地采用了虚拟化技术，包括虚拟服务器、虚拟存储和虚拟网络等，虚拟化安全成为了 IaaS 平台面临的最大安全风险。虚拟化安全可以归结成以下两个方面。

1. 虚拟化软件安全

该软件层部署于物理裸机上，用于提供创建、运行和销毁虚拟服务器的服务。为了保证 IaaS 层的安全，云服务提供商需要建立完善的安全措施，用于限制对 Hypervisor 和其他形式的虚拟化层的物理和逻辑访问。在 IaaS 层的服务中，用户不能直接接入虚拟化软件层，该层由云服务提供商来负责操作和管理。

2. 虚拟服务器安全

虚拟服务器和客户端面临着各种主机安全威胁，例如接入和管理主机的密钥可能被盗、在没有完备安全措施的服务标准端口侦听和劫持访问账户等，要应对这些主机安全威胁就需要在虚拟服务器中采取如下措施：

（1）选择具有 TPM（可信平台模块）的虚拟服务器，以保证服务器具有真正的可信性。

（2）为每台虚拟服务器分配独立的硬盘分区，以便实现服务器的逻辑隔离。

（3）虚拟服务器之间还应采用 VLAN 和划分不同 IP 网段的方式进行逻辑隔离，需要通信的虚拟服务器间的网络连接可以采用 VPN 技术进行安全保证。

（4）对虚拟服务器进行有计划的备份，可采用完整、增量或差量的备份方式。

5.5.3 PaaS 层安全问题及措施

PaaS 层位于云计算平台的中间层，它依靠 IaaS 层提供的基础资源，为 SaaS 层提供应用平台，起到了承上启下的作用。

PaaS 层的核心技术是分布式处理，主要解决大规模服务器群的协同工作的问题。要提供 PaaS 层的云计算服务，首先需要在云计算数据中心架设分布式的处理平台，包括分布式的文件系统、分布式计算及分布式数据库等；其次，需要对分布式处理平台进行封装，提供开发环境（Software Development Kit）、访问 API 接口及开发工具库等。PaaS 层面临的安全威胁主要体现在如下两个方面。

1. 分布式文件及数据库安全

由于云计算数据中心的分布式文件系统和分布式数据库系统一般都构建在大规模廉价的服务器集群上，从而使得分布式文件系统和数据库系统面临诸多的安全挑战。例如，由于容错问题解决不好而导致的服务器的失效现象经常出现；由于服务器增减频繁而需要解决服务器的动态扩展问题；需要提供对于海量数据的存储及快速检索能力；需要解决由于多用户同时访问而带来的并发控制和存取效率等问题。

为了提高分布式文件系统的可靠性以避免服务器失效，可行的做法是采取冗余存储的方式，在系统中保存每份文件或数据的多个备份。冗余存储方式可以解决数据的可靠性问题，但也带来了数据一致性的问题，因为云计算中文件或数据存储在多个不同的节点上，冗余存储方式下在对文件或数据进行修改时，必须确保对所有的存储副本都进行了修改，这就必须有分布式的同步机制用于对并发操作进行控制。这些技术的复杂性可能给数据的可靠和安全带来挑战。

2. 用户接口及应用安全

基于来自客户端的服务都可能带有恶意的考虑，如果 PaaS 层暴露的接口过多，都可能会给攻击者带来更多的攻击机会，比如抢占 CPU 的时间、内存空间及其他资源，也可能会攻击其他的用户，甚至可能会攻击云服务的底层平台等。因此，认证用户的可靠性是 PaaS 层面临的安全问题。

基于 PaaS 平台开发的软件都将部署在该平台上，PaaS 提供商必须能保证程序的可靠运行，尤其是需要保证不同应用之间的相互隔离。这一点与 SaaS 模式下遇到的安全挑战是类似的。另一方面，目前 PaaS 对于底层资源的调度及分配都是采用“尽力而为”的机制，如果在一个平台上运行多个应用，就不可避免地存在资源分配和优先级配置的问题，多个应用的资源调配就需要借助 IaaS 层的虚拟化机制来实现，由此带来的安全问题，由 IaaS 层的安全技术来保证。

5.5.4 SaaS 层安全问题及措施

传统软件都部署在客户自己物理机或是租用的数据中心，由于仅服务于特定的用户，所以其安全控制相对简单。然而在云计算模式下，成千上万的客户可能共享同一软件平台，如何保证客户之间的数据和应用的安全是一个巨大挑战，这一问题属于 SaaS 层的安全问题。多租户技术使得大量的用户共享同一软件资源，每个用户都是按需使用软件资源。多租户技术提供对软件服务的客户化配置，而不影响其他用户的使用。目前解决 SaaS 模式安全问题主要是依赖多租户技术，然而这一技术也存在数据隔离、客户化配置等多方面的问题需要考察。

1. 数据隔离

数据隔离是指在同一个 SaaS 系统中，不同租户的数据是被隔离存储的，系统对数据的处理不会相互干扰。目前，实现多租户之间的数据隔离有 3 种技术：

（1）为每一个租户提供单独的数据库，以此保证不同租户之间数据的充分隔离。此技术的问题是成本和开销都比较大。

（2）多个租户的数据保存在同一数据库中，但是采用不同的模式，这种方式在一定程度上可以减少数据库的访问成本和操作难度，同时带来的问题是影响了数据隔离效果及安全性。

（3）多个租户的数据保存在同一数据库的同一张数据表中，通过租户标识字段来区分数据的拥有者，这种方式成本最低，但安全性和隔离性最差。

2. 客户化配置

客户化配置是指 SaaS 应用允许不同的租户对同一软件平台进行个性化的定制。在传统应用中，每一个用户都拥有自主的应用实例，可以进行定制化的开发。但在云计算的多租户形式下，多个租户共享的是同一个应用实例，如果某一个租户对应用的配置更改会对该平台的所有租户产生影响，也将带来更多的应用问题。因此，如何使得不同的租户对同一实例实现独立的客户化配置，是 SaaS 模式需要面对的一个安全挑战。

除了上述安全风险以外，多租户技术还面临着架构拓展和性能定制等安全方面的挑战。为了解决这些问题，有人建议采用虚拟化技术。面对大量用户使用统一应用时，如果采用虚拟化技术把每一个用户的应用都做成一个单独的虚拟机，就可能需要成千上万的虚拟机，这使得系统的管理难度和复杂性都会增加，这方面还存在大量需要研究的课题。

5.5.5 云计算安全关键技术

虽然在云计算的各层所面临的安全风险不同，可以采取的安全技术也不尽相同，然而有几大技术却是具有共性的，下面说明在云计算环境下几种主要的安全

技术。

1. 虚拟化安全技术

虚拟化是云计算的核心技术之一，虚拟化技术的应用加快了传统应用部署的速度，使得应用的兼容性和服务的可用性得到提高。与此同时，虚拟化自身也存在诸多风险和威胁，虚拟化安全成为云计算面临的主要安全威胁之一。概括起来，目前可用于虚拟化安全的主要措施包括可信平台虚拟机、虚拟机隔离、虚拟机信息流的控制、虚拟机监控及虚拟网络的接入控制等。

2. 数据安全

由于用户对数据的不可控制性，云计算环境下的数据安全成为客户最关注的问题，所面临的安全挑战包括：

(1) 数据存放位置。必须保证所有数据（包括所有副本和备份）存储在合同、服务水平协议和法规所允许的物理位置。

(2) 不同客户数据的混合。数据尤其是保密和敏感数据在其被使用、存储或传输的过程中，不能与其他客户的数据混合。数据的混合将会带来数据安全方面的更多安全挑战。

(3) 数据备份和恢复重建。要保证数据的可用性，必须建立有效的云数据备份和恢复计划，有效地避免数据丢失和意外的数据破坏。

(4) 数据删除或持久性。必须具有一种可信技术，使得用户可以全面和有效地定位及销毁云计算的数据，确保数据已被完全地消除，并且无法恢复。

针对数据安全问题，通常有数据隔离、数据加密、数据切分、数据屏蔽以及数据删除等技术进行解决。其中，数据隔离和保密等涉及用户隐私保护的技术仍然是这方面的研究重点。

3. 云资源访问控制

在云计算中，不同云应用可以属于不同的安全管理域，每个安全域都负责管理着本地资源和用户。当用户跨域访问资源时，需要在域边界进行认证服务，对于访问共享资源的用户需要进行统一的身份认证。

传统模式下的身份认证和访问管理技术已经比较成熟，然而云计算模式下，服务商（IaaS、PaaS、SaaS）所支持的标准并不健全，难以满足企业对监测管理、隐私性以及数据保护方面的需求。目前，云计算环境下的身份认证和访问控制管理方面还存在很多不足，需要更深入地研究和探索。

通过以上安全技术，可以提供满足需求的云安全服务，为各类云应用提供共性的信息安全服务。云安全服务属于云基础软件服务层，其中比较典型的几类云安全服务包括加密与密钥管理、身份识别与访问控制、灾备与业务连续性、数据隔离技术及虚拟化安全技术等，这些技术对云计算的几种模式都会产生影响。



第6章

面向省市的社区矫正信息系统

6.1 概述

为了更好地促进社区矫正工作向合理化、人性化、智能化、效率化方向发展，推动社区矫正工作管理的进步，建设社区矫正信息系统对于促进司法行政信息化建设战略具有重要意义。

社区矫正管理系统为省司法厅、市司法局、区县司法局及乡镇街道司法所各级司法行政工作人员提供社区矫正的全业务流程信息化管理。系统围绕社区服刑人员、监管人、责任人和社区矫正工作等几条管理主线，实现对社区服刑人员从衔接到解除全过程的动态管理和量化管理，重点实现对社区矫正刑罚执行活动过程的规范控制和程序化管理，其主要工作流程包括：矫正衔接、矫正执行、管理监督、考核奖惩、矫正解除。

通过社区矫正管理系统，能够落实社区矫正的各项工作制度，简化社区矫正管理程序，降低管理成本，提高工作效率。利用先进的信息技术，实现“区域监管、信息交互、警示告知、考核管理”等四大功能，可以随时随地了解矫正对象的位置监管，进行高效的信息交互，解决矫正对象越界告警、到期警示等监督难题。

6.1.1 系统设计原则

系统设计中应该遵循如下原则。

1. 标准化原则

规范性、标准化是一个大型系统建设的基础，也是系统与其他系统兼容和进一

步扩充的根本保证。因此，对于社区矫正系统来说，数据的规范性和标准化工作是极其重要的，是系统开放性和数据共享的要求。在系统建设之前应有明确统一的数据采集规范和质量标准。系统建设需要遵循国家规范标准和司法行业规范标准。

2. 因地制宜原则

充分借鉴其他行业经验，结合社区矫正管理的实际情况，确定系统功能，注重实效，力求在省市及基层社区矫正信息化建设中有所创新。

3. 先进性原则

在系统的总体设计上，应该借鉴各类系统的成功经验，注重吸取同类系统的建设教训；在技术上，要采用国际上先进的且成熟的技术，使得设计更加合理、更为先进。同时，也要充分考虑社区矫正系统的现状和特点，在注重系统实用性的前提下，尽可能采用先进的计算机软、硬件环境；在软件的开发思想上，严格按照软件工程的标准和面向对象的理论来设计，保证系统的先进性。

4. 安全性原则

由于整个系统所涉及的数据大多属于内部资料，这些数据的安全性至关重要，因此，系统应遵循安全性的原则。安全性问题主要包括以下三种情况：一是防止外部非法用户访问网络；二是防止内部合法用户的越权访问；三是意外的数据损坏。为了提高系统的安全性，在设计时需要考虑两个系统的安全问题：一是网络建设中部署安全设备，并考虑冗余策略；二是对系统内部不同的用户、不同的部门分别赋予不同的权限等。

5. 可扩充性原则

面对信息技术的高速发展，系统的计算机设备、网络设备和应用软件都应具备良好的系统扩充性。随着网络技术的不断发展，主干网络设备应该能够实现平滑升级，因而在社区社区矫正系统设计的过程中应保证系统结构模块化和系统软硬件平台的可扩展性。

6. 稳定性原则

稳定性一般指系统的正确性和健壮性。系统是在网络环境下运行的，因而存在系统管理数据量较大，数据使用并发性强等问题，这些问题对系统的设计提出了更高的要求。为了解决上述问题，一方面，系统必须在提交前进行反复测试，并在系统正式投入运行前实行试运行机制，这样可以把出错率降到最低，以保证系统的正常运转；另一方面，系统必须有足够的健壮性，在发生意外和灾难性破坏的情况下，不仅能够给出错误提示，而且还可以很好地处理这些错误，使系统能够及时恢复原始状态，减少不必要的损失。

6.1.2 系统工作流程

面向省市级的社区矫正工作流程如图 6-1 所示，包括省司法厅、市司法局、区县司法局及基层司法所。省司法厅主要实现省厅的监督、指导职能，具体包括通知公告管理、监督考核管理、统计分析及报表、数据交换等。

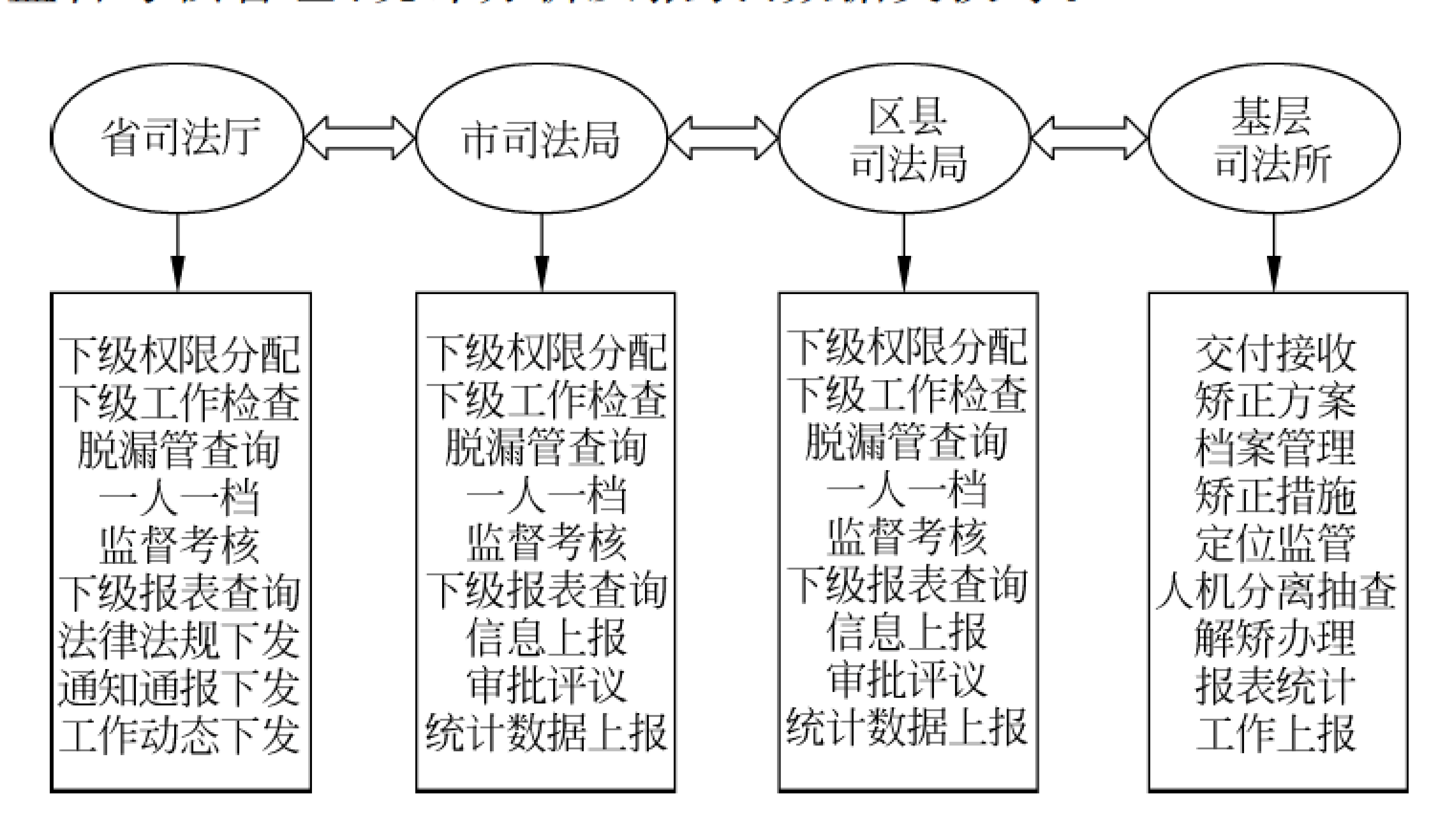


图 6-1　省市社区矫正工作流程

市级司法局偏重于具体的业务，负责矫正对象信息录入、监管、考核等工作，具体包括：

(1) 人员监管——实现识别、定位及行迹监测等功能；

(2) 档案管理——对矫正对象的基本资料、社会关系、帮教小组等信息进行录入与维护；

(3) 考核管理——实现矫正对象的思想汇报情况、公益劳动登记、学习教育、请假登记等功能。

区县司法局和基层司法所与市级司法局类似，偏重具体的业务，如交付接收、矫正方案、档案管理、矫正措施、定位监管、解矫办理及工作上报等。

6.1.3 系统总体逻辑结构

根据社区矫正工作业务的实际需求，系统逻辑结构如图 6-2 所示。

从如图 6-2 所示的逻辑结构可以看出，社区矫正管理系统由四部分组成。

1. 数据层

系统采用集中的数据管理方式，通过建立面向省市区县社区矫正人员定位业务的数据模型，将空间数据、业务数据和系统管理数据进行标准化和规范化管理，建立各类数据内部及其数据间的关系。为在统一框架下实现省市社区矫正人员定位信息数据的关联、功能调用和数据交换共享奠定基础。

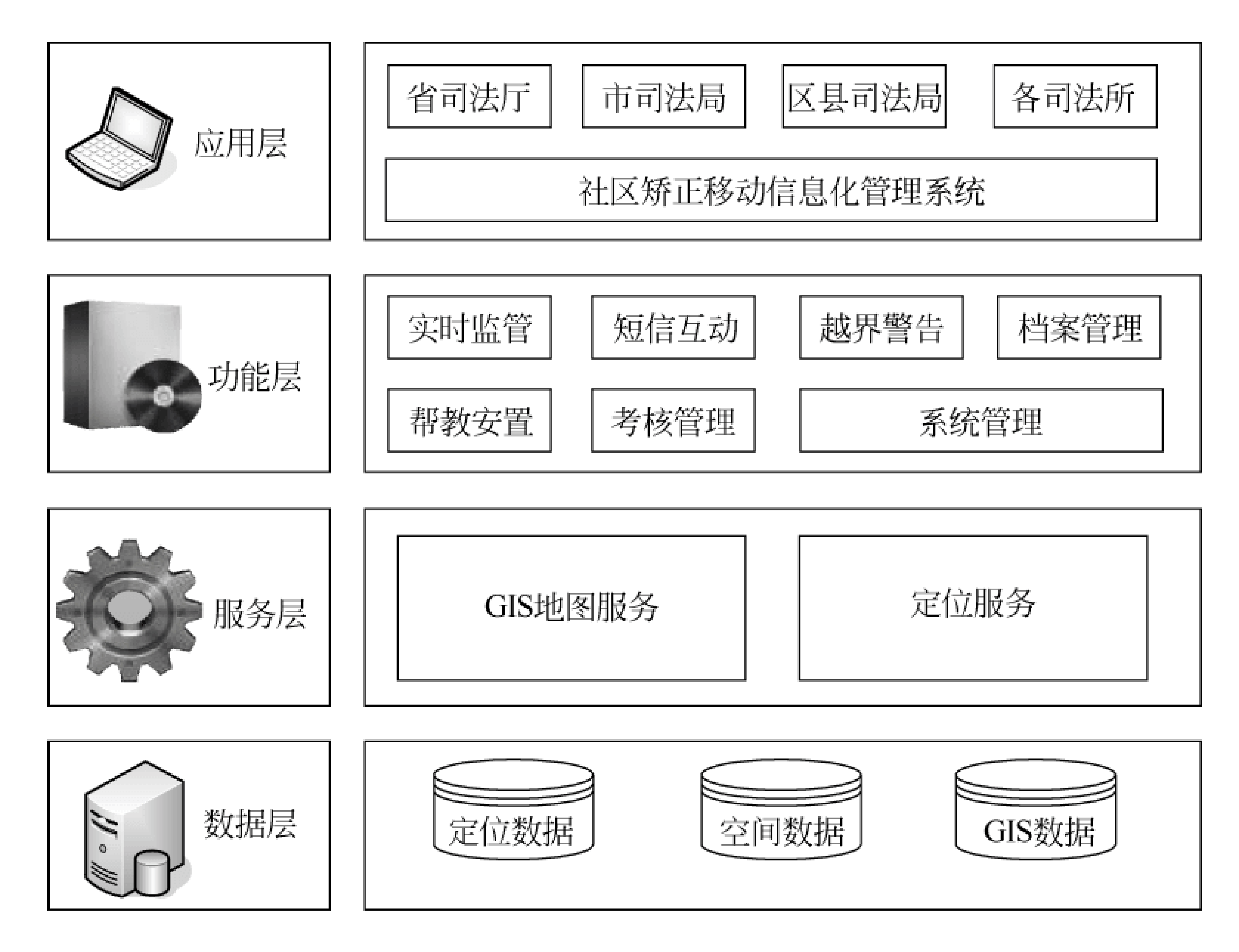


图 6-2 系统总体逻辑结构

2. 服务层

服务层是整个系统的骨干，主要包括 GPS 服务、LBS 服务、GIS 应用服务、数据管理服务等。服务层组件向各应用提供服务。

3. 功能层

这些应用模块构成社区矫正移动信息化管理系统的基础部分，应用模块主要包括基础地理信息管理、实时监管、短信互动、网上办公、档案管理、帮教安置、系统管理等模块。这些应用模块通过各种不同的组件进行构建，应用模块之间不仅可以共享其他应用模块的组件，也可以共享本应用模块专用的组件，从而实现不同业务子系统的功能调用和数据访问，形成一个完整的业务子系统。这样形成的业务子系统虽共享组件，但功能却相对独立，因而可以保障社区矫正信息化管理系统各项业务功能的顺利实现。

4. 应用层

指基于数据和服务构建的、实现具体业务的应用系统，为最终的司法行政用户提供服务。

6.1.4 社区矫正信息化平台建设

针对传统社区矫正单一文书档案的管理方式，建立一个社区矫正信息化平台，该平台用于完善社区矫正工作制度，使得档案资料的管理方式从文书档案管理转

变成电子档案管理，在简化工作手续的同时规范工作流程。

根据社区矫正管理的需要，信息化平台的建设既可以采用 B/S 结构，也可以采用 C/S 结构，该平台为矫正工作者提供可视化的、友好的、可交互式的工作方式。

为了保证系统安全性，信息化平台将信息化服务和数据库服务的服务器建设在司法局的网络中，且仅开放需要访问的端口。在此基础上对使用信息化平台的人员实行多级密码管理及多级权限管理机制，这样可以从物理和逻辑上为系统提供足够的安全性保证。

社区矫正信息化平台提供的功能包括建立矫正档案、修改矫正档案、删除矫正档案、多关键字档案查询、档案模糊查询、权限管理、定位管理、短信管理、报警管理、网上办公、安置帮教、统计查询等诸多功能。具体来说，主要包括以下五大类功能：

1. 人员定位管理

主要对矫正人员进行定位、短信、报警等管理。

（1）矫正定位。

（2）短信互动。

（3）报警管理。

（4）语音比对。

2. 档案工作管理

主要记录矫正人员的电子档案信息、每月的矫正工作信息及进行数据报表统计。

（1）入矫管理。

（2）解矫管理。

（3）工作管理。

（4）报表管理。

3. 网上办公管理

主要用于司法所工作人员上报矫正人员奖惩信息和请假信息，区县司法局通过系统进行网上审批，减少审批流程。

（1）审批管理。

（2）发布管理。

（3）消息管理。

4. 帮教安置管理

主要记录刑释人员档案信息及报表统计。

（1）帮教安置管理。

（2）报表管理。

5. 系统管理

主要对系统的组织机构、用户权限、登录用户名密码等进行管理。

（1）组织机构管理。

（2）基地、社区管理。

（3）用户管理。

（4）权限管理。

6.2　省市级社区矫正信息系统中心

6.2.1　网络连接结构

省市级社区矫正系统覆盖范围涉及全省（自治区、直辖市），实现全省（区、市）统一管理、统一发布，这样不仅可以增强信息流转的时效性，同时也便于政策、方针的统一宣传学习和贯彻落实。采用全省数据集中存储的方式，有利于实施应急备份方案，数据安全级别较高。此外，采用统一模式、统一标准建设该矫正系统，使得系统的部署实施工作更为简单易行，地级市、区县的数据复制也更加方便。

省市级社区矫正系统需要依托政法网组建广域网，或者建立自己的专网，连接省司法厅、各地市及县乡司法所，建设符合国家电子政务统一标准的网络平台。该平台在符合司法行政系统信息安全和网络安全要求的基础上，不仅满足社区矫正系统网络传输和业务处理功能的需求，而且具有高效的信息处理能力和可靠的信息处理安全技术手段。网络结构分为省级信息中心、市级网络、县区网络及乡镇网络四级，其网络拓扑结构如图 6-3 所示。

由于司法行政系统存在大量涉密信息，因而该系统属于涉密系统。在规划和建设计算机信息系统时，需要做到严格遵守国家有关保密制度，同步规划落实相应的保密措施，以确保系统的安全。

1. 信息安全基础设施

（1）数字证书体系；

（2）网络病毒防治服务体系；

（3）数据备份和容灾体系；

（4）应急响应和支援体系。

2. 网络安全防护体系

（1）物理层安全，保证计算机信息系统各种设备的物理安全是整个计算机信息系统安全的前提。物理安全是指保护计算机网络设备和其他媒体免遭地震、水

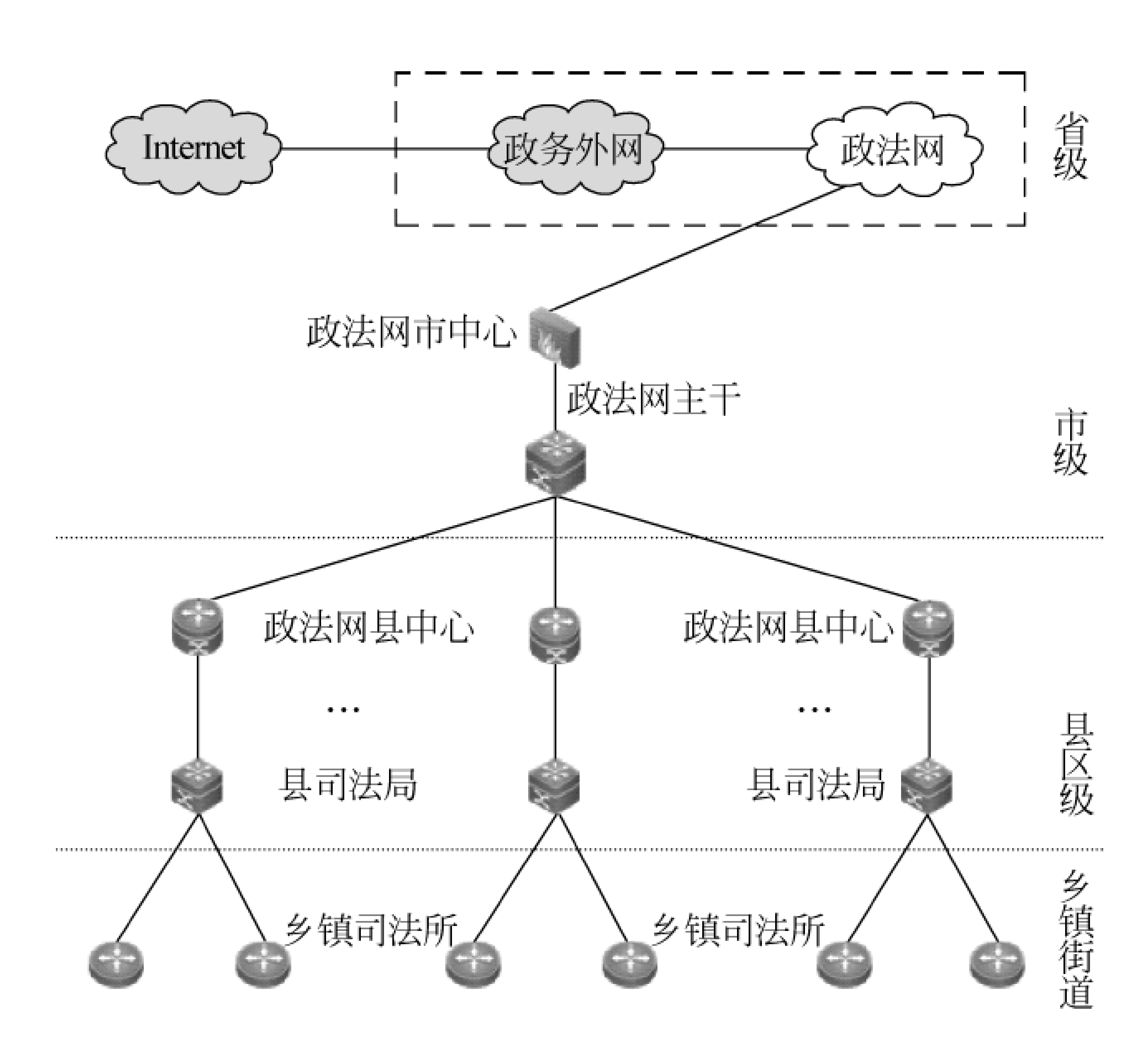


图 6-3 社区矫正系统网络拓扑结构

灾、火灾等环境事故、人为操作失误及各种计算机犯罪行为导致的破坏，主要包括机房环境、物理链路安全等。

（2）网络层安全，包括防火墙、入侵检测、网络加密、网络漏洞扫描等。

（3）系统层安全，包括主机访问控制，系统漏洞扫描、病毒防范等。

（4）应用层安全，包括身份认证、内容过滤等。

（5）管理层安全，包括安全管理策略、安全审计、统一安全管理等。

3. 常用防范措施

1）防火墙部署

在网络边界部署防火墙，成为政务外网与互联网之间的安全屏障。其中服务器连接在防火墙的 DMZ(Demilitarized Zone，非军事区)区，与外界进行隔离，内网口连接政务外网交换机，外网口通过路由器与外界连接。这样，外界用户只能访问到对外公开的一些服务(如 WWW、MAIL、FTP 等)，既保护内网资源不被外部非授权用户非法访问或破坏，也可以阻止内部用户对外部不良资源的滥用，并能够对发生在网络中的安全事件进行跟踪和审计。在防火墙设置上可按以下原则配置以提高网络安全性：

（1）根据网络安全策略和安全目标，规划设置正确的安全过滤规则，规则审核 IP 数据包的内容包括协议、端口、源地址、目的地址、流向等项目，严格禁止外界对内部网不必要的、非法的访问。总体上遵从“不被允许的服务就是被禁止”的原则。

（2）将防火墙配置为：过滤掉以内部网络地址进入路由器的 IP 包，这样可以

防范源地址假冒和源路由类型的攻击；过滤掉以非法 IP 地址离开内部网络的 IP 包，防止内部网络发起的对外攻击。

（3）在防火墙上建立内网计算机的 IP 地址和 MAC 地址的对应表，防止 IP 地址被盗用。

（4）定期查看防火墙访问日志，及时发现攻击行为和不良上网记录。

（5）允许通过配置网卡对防火墙设置，提高防火墙管理安全性。

2）入侵检测系统部署

入侵检测系统一般集入侵检测、网络管理和网络监视功能于一身，能实时捕获内外网之间传输的所有数据，利用内置的攻击特征库，使用模式匹配和智能分析的方法，检测网络上发生的入侵行为和异常现象，并在数据库中记录有关事件，作为网络管理员事后分析的依据；如果情况严重，入侵检测系统可以发出实时报警，使得网络管理员能够及时采取应对措施。

3）漏洞扫描系统

利用先进的漏洞扫描系统定期对工作站、服务器、交换机等进行安全检查，并根据检查结果向系统管理员提供详细可靠的安全性分析报告，为提高网络安全整体水平提供重要依据。

4）网管系统

该系统能够管理各种品牌的网络设备和服务器的真实拓扑、设备动态真实背板和自动报表等，大大减少网管员工作的复杂度和烦琐的工作。

5）网络版杀毒产品部署

通过部署网络版杀毒产品，实现在整个局域网内杜绝病毒感染、传播和发作的效果。在整个网络内，对可能感染和传播病毒的地方，均采取相应的防病毒措施，同时为了能够有效、快捷地实施和管理整个网络的防病毒体系，为系统安装的网络版杀毒产品应具有远程安装、智能升级、远程报警、集中管理、分布查杀等多种功能。

4. 安全管理与标准法规

为了制定安全管理与标准法规，不仅要加强对人员的管理和完善相应的规章制度，以便形成完整、规范的安全管理体系，还要严格遵守《计算机信息网络国际联网安全保护管理办法》、《计算机病毒防治管理办法》、《国家信息化领导小组关于加强信息安全保障工作的意见》，并遵照执行《计算机信息系统安全保护等级划分准则》（GB17859—1999）、《信息安全管理实施细则》（BS7799—1：1999）、《信息安全管理体系规范》（BS7799—2：1999）等技术规范。

6.2.2 中心网络结构

1. 数据中心平台

数据中心不仅要能满足应用系统建设的基本要求，还要充分考虑到司法行政信息化建设未来发展的需求，且具备一定的扩展性。硬件平台建设规划在保证技术先进性、现实性和发展性的同时，最大限度地保护既有投资，减少对系统的维护和未来开发的成本，实现技术上、经济上的可持续发展。完整的数据中心基础架构如图 6-4 所示。

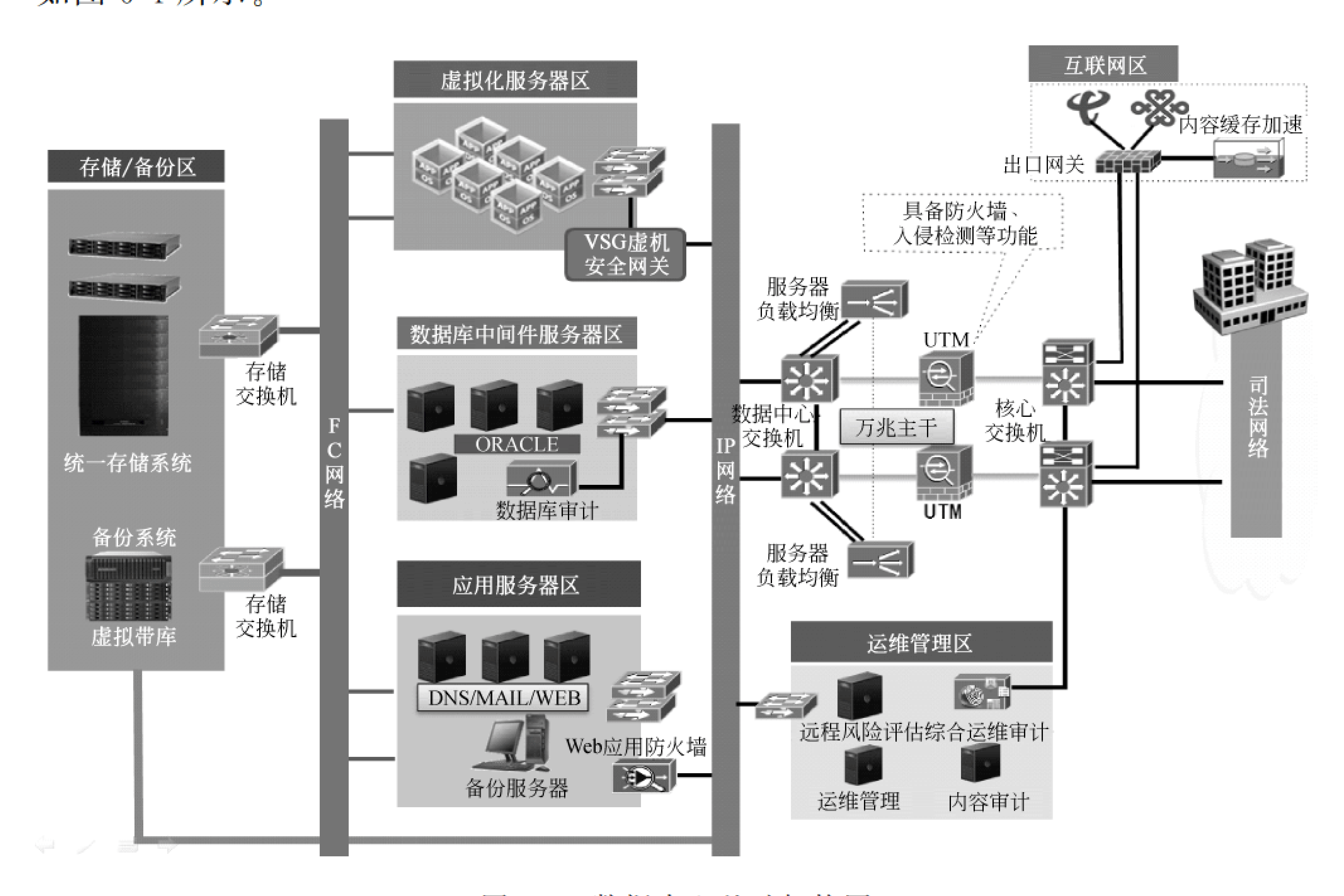


图 6-4 数据中心基础架构图

2. 服务器系统

数据中心承载着所有的司法行政信息化应用系统，不同应用系统的部署对硬件资源的要求也各不相同。数据中心服务器系统的建设除了需要有高性能方面的保障，同时也要能确保系统的高稳定性；不仅可以支持多种灵活的系统集群、迁移等方面的技术，还要支持简单有效的统一配置、管理手段。

服务器系统建设建立在“整合”的概念和技术基础之上，具备下列特点：有限的设备数量、共享资源、策略化的系统自动管理和部署。在服务器类型的选择上有两种方式：一种是传统的机架式服务器，另一种是基于服务器整合的刀片中心。

从安全、稳定角度考虑，数据库服务器建议采用机架式小型机，可以满足未来5～10年内的性能需求；同时该设备支持逻辑分区和虚拟化功能，可以将一些重要的应用系统整合到同一设备的不同分区中部署，并可实现HA(High Available，高可用)集群。

Web服务器和应用服务器建议采用刀片服务器或高性能机架式服务器，通过虚拟化整合的方式提高硬件资源的利用率。

同时，还可以考虑配置服务器负载均衡器和智能DNS系统，提高省、市、县、社区各层次用户的访问速度。

3. 存储系统

作为数据中心最重要的数据存储平台，不但要实现系统高可用性、高性能、连接与支持能力，还要考虑系统的扩展能力和兼容性；同时，对现有存储系统进行整合，实现核心关键业务系统同城数据中心双活，非关键业务系统实现RTO(Recovery Time Object，恢复时间目标)小于24小时。具体建设内容有：

1) 数据存储系统

在数据中心部署一套虚拟化存储引擎，使用两台存储设备。当数据中心任何存储设备出现故障后，可以确保RPO(Recovery Point Object，恢复点目标)＝0，RTO≈0。虚拟化存储引擎可实现同型号存储的同步镜像，还可实现不同品牌型号存储设备的同步镜像。

2) 备份与容灾系统

建设数据库备份，可以通过软件结合虚拟带库实现对存储上数据库数据的备份，也可以通过SAN(Storage Area Network，存储区域网络)网络对各个业务数据库进行在线备份，且备份作业不占用生产网络资源。其中虚拟带库具有NAS(Network Attacker Storage，网络附加存储)归档功能，可以对档案系统等类型业务进行集中的文件存储，也可以通过强大的数据消重技术降低存储空间用量。当在异地部署一台虚拟带库时，可通过以太网进行两台设备之间的数据复制，实现关键数据备份的异地容灾。

该存储方案具有如下特点：

(1) 解决存储单点故障。

实现生产数据中心存储双活，可以满足并实现核心业务系统等关键业务数据不丢、业务不停。

(2) 无中断数据迁移。

可以满足系统升级、设备更新换代、机房搬迁的需要。当数据迁移需长时间业务停顿时，可通过虚拟存储引擎实现数据中心内部及跨数据中心无中断数据迁移。

（3）提高资源利用率、简化管理。

通过虚拟化技术实现异构存储统一管理及优化整合，同时提高资源使用效率，降低管理复杂性。

（4）存储双活。

存储故障时无须人工干预即可实现业务连续，同时采用分布式缓存一致性技术保证随处访问，提高系统的灵活性和可用性。

一旦数据中心某个存储阵列出现故障导致停机，采用双活数据中心解决方案可以确保业务数据不丢、业务不停，无须手工切换即可保证业务系统服务的连续性。

6.3　基层社区矫正管理软件系统

6.3.1　社区矫正管理信息系统分析

随着社会信息化程度的提高，高科技与人们的生活和工作也越来越密切，社区矫正的管理也从静态封闭式的环境走向了动态环境，社会服刑人员的控制难度随之增大。在这种情况下，急需社区矫正信息化建设。为了规范社区矫正信息化建设，司法部出台了《社区矫正管理信息系统技术规范》和《社区矫正人员定位系统技术规范》，规定了社区矫正管理信息系统的基本功能、数据规范、编码规则、数据交换以及系统安全等内容，是研发社区矫正相关业务信息系统的依据。

下面将依据上述规范进行社区矫正管理信息系统进行分析论证。

1. 社区矫正管理信息系统功能概述

根据司法部《社区矫正管理信息系统技术规范》要求，结合社区矫正工作的具体情况，社区矫正管理信息系统主要包括位置监控系统、警示告知系统、档案管理系统、流程管理系统、文书管理系统、日常管理系统、权限管理以及工作机制等主要功能模块。每个模块的具体功能如下：

1）位置监控系统

通过无线网络和定位技术，实现矫正对象定位监控、位置查询、运动轨迹回放等功能，方便管理员对矫正对象进行监督和管理，确保矫正对象在所限制区域内，没有违法违规等行为。

2）警示告知系统

越界告警：出现矫正对象越界、关机等情况时，系统自动向监控平台发出报警信号，并以短信形式警告矫正对象。

到期警示：矫正对象矫正期满前，系统会提示矫正办公室进行期满鉴定，办理

解除矫正手续。

警示记录：记录警示的发生类型、时间日期等内容，通过日志形式保存在本地方便管理员的查询和存档。

3）档案管理系统

建立每个矫正对象的数字档案，详细记录矫正对象的各种信息，包括他们从宣判到解矫全过程的信息，并且建立多维查询系统，方便、简单、快捷地查询其信息。

4）流程管理系统

负责矫正管理系统各项流程的规范介绍、创建、监督、修改、删除等管理工作。每项矫正事务有相应的处理流程，系统按照既定的流程执行操作，并且可以查看每项事务流程的处理进度和处理结果。

5）文书管理系统

社区矫正过程中各个部门之间应按规定的文书格式信息进行交互，这些文书包括调查评估意见书、社区矫正人员基本信息表、社区矫正人员报到情况通知单、社区矫正宣告书等二十多个文书，文书管理系统可对这些文书进行安全有效的管理。

6）日常管理系统

管理社区矫正对象的集中学习、公益劳动、心理矫正、生活、思想汇报等事项。并通过监控、采集、记录、管理矫正对象表现信息，包括心理特征及矫正方案、违规违纪记录、奖励记录、月考核、季考核、年度评定情况等，作为考核依据。

7）权限管理

管理矫正管理人员的操作权限，如用户口令、密码，读、写、修改数据权限等，同时对系统进行维护和改善。

8）工作机制

包括政府工作机制、工作队伍和基地建设等项目，主要用于关联相关职能机构，建立相应的司法行政体系和工作队伍。

社区矫正管理信息系统的功能模块结构如图 6-5 所示。

2. 社区矫正管理信息系统需求分析

1）系统用户介绍

根据相关规定，社区矫正管理工作的组成人员包括司法所人员、社会工作者、志愿者、村委会人员、所在单位人员、就读学校人员、家庭成员或监护人以及保证人等。此外每个社区矫正管理子系统还包括系统管理员，用来管理工作机制和系统维护，其用例如图 6-6 所示。

系统用户的职责描述如表 6-1 所示。

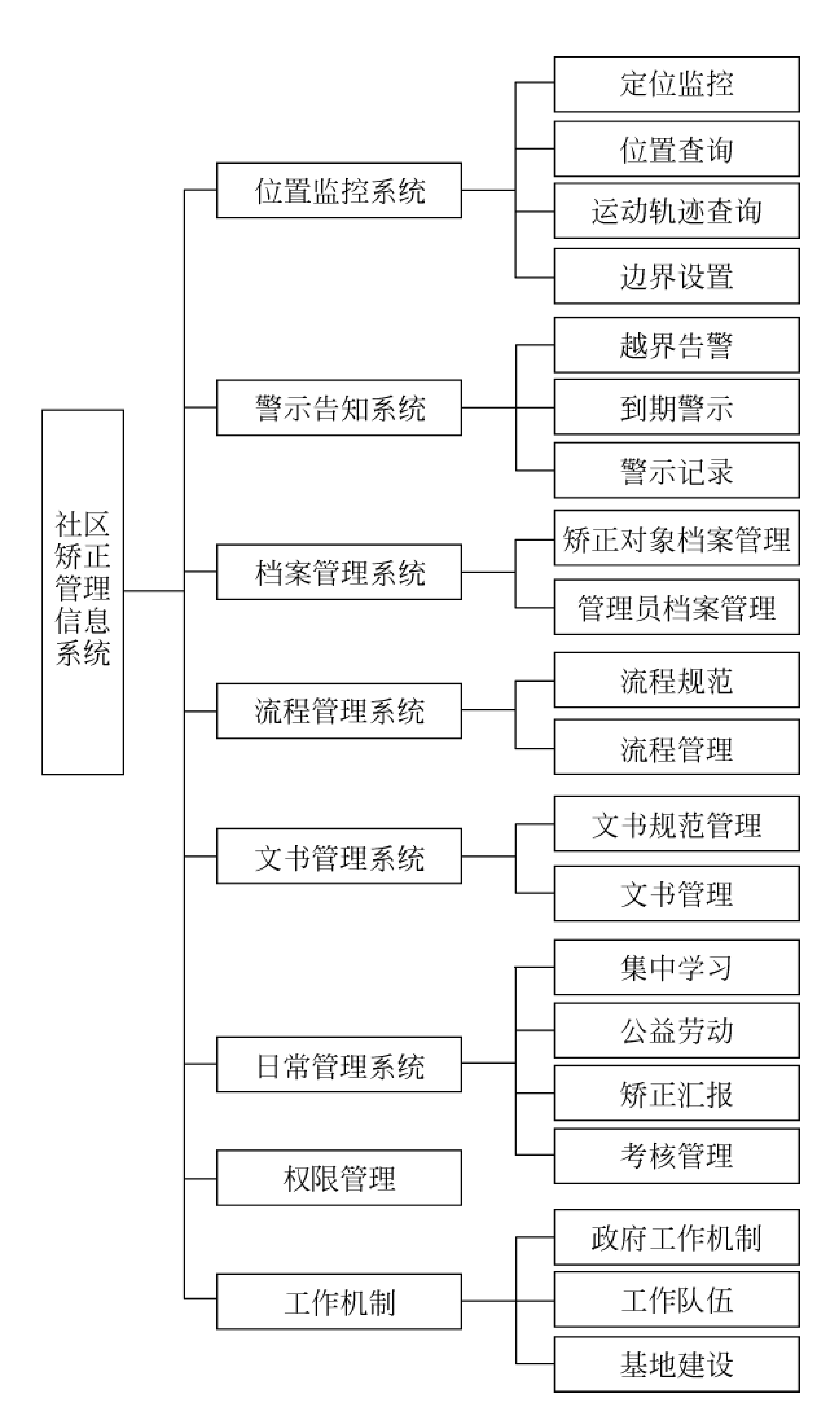


图 6-5　系统功能模块结构

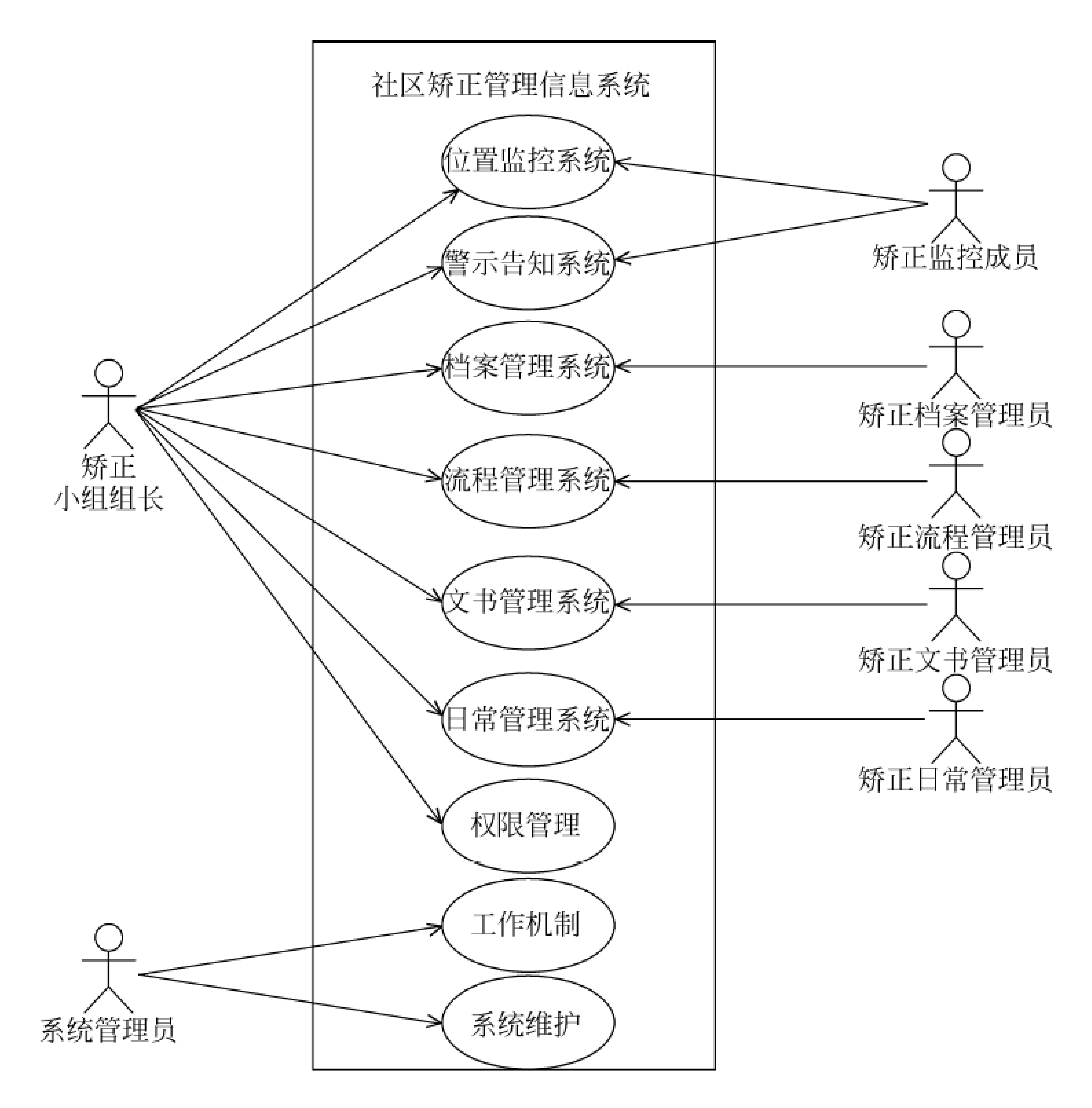


图 6-6　系统用例图

表 6-1　系统用户的职责描述

系统用户	操作范围
系统管理员	负责系统的维护，具有系统的最高权限，由司法所内部技术人员担任，系统的升级及更改需通过领导审批通过
矫正小组组长	负责矫正工作的任务分配及矫正监督，是除系统管理员外的最高权限者，通过系统能查看矫正对象的所有信息及矫正情况，还负责文书审批等工作
矫正监控成员	负责矫正监督、警告处理等监管工作，具有位置查询、轨迹查询、边界设置、警示汇报等职能，同时负责矫正中出现的违法、脱逃等事务。由司法局工作人员担任
矫正档案管理员	负责矫正对象档案管理，包括矫正工作中有关矫正对象的各项信息及信息的变更等，由司法局工作人员担任
矫正流程管理员	负责矫正工作审批流程规范和流程管理，文书在审批过程中确保安全和可靠性。由司法局工作人员担任
矫正文书管理员	负责矫正文书管理工作，包括矫正文书的下发、上报等工作，按照矫正流程负责将文书提交给各个上级主管及部门，由司法局工作人员担任
矫正日常管理员	负责矫正的日常事务，包括集中学习、公益劳动、矫正工作汇报以及考核管理。由社会工作者、志愿者、村委会人员、所在单位人员、就读学校人员、家庭成员或监护人以及保证人等监督实施

2）系统适用人群

社区矫正适用范围是户籍地和经常居住地在本地，或者户籍地不在本地，本人及家庭成员在本地定居，且长期在本地工作、学习和生活的下列五种罪犯：

（1）被判处管制的；

（2）被宣告缓刑的；

（3）被裁定假释的；

（4）被暂予监外执行的；

（5）被剥夺政治权利并在社会上服刑的。

3）系统用例模型

用例模型的设计用来描述“用户、需求、系统功能单元”三者之间的关系，展示了一个外部用户能够观察到的系统功能模型图。开发软件就好像建造房子一样，都需要先设计好模型，越复杂的结构越需要分析设计好模型，模型设计的好坏直接影响最后的结果。因此，用例是划分系统功能强有力的工具，对于分析系统结构和功能描述必不可少。前面介绍了社区矫正管理信息系统的基本功能，下面将以用例图分析每个系统模块包含的具体功能。

（1）位置监控系统用例模型，如图 6-7 所示。

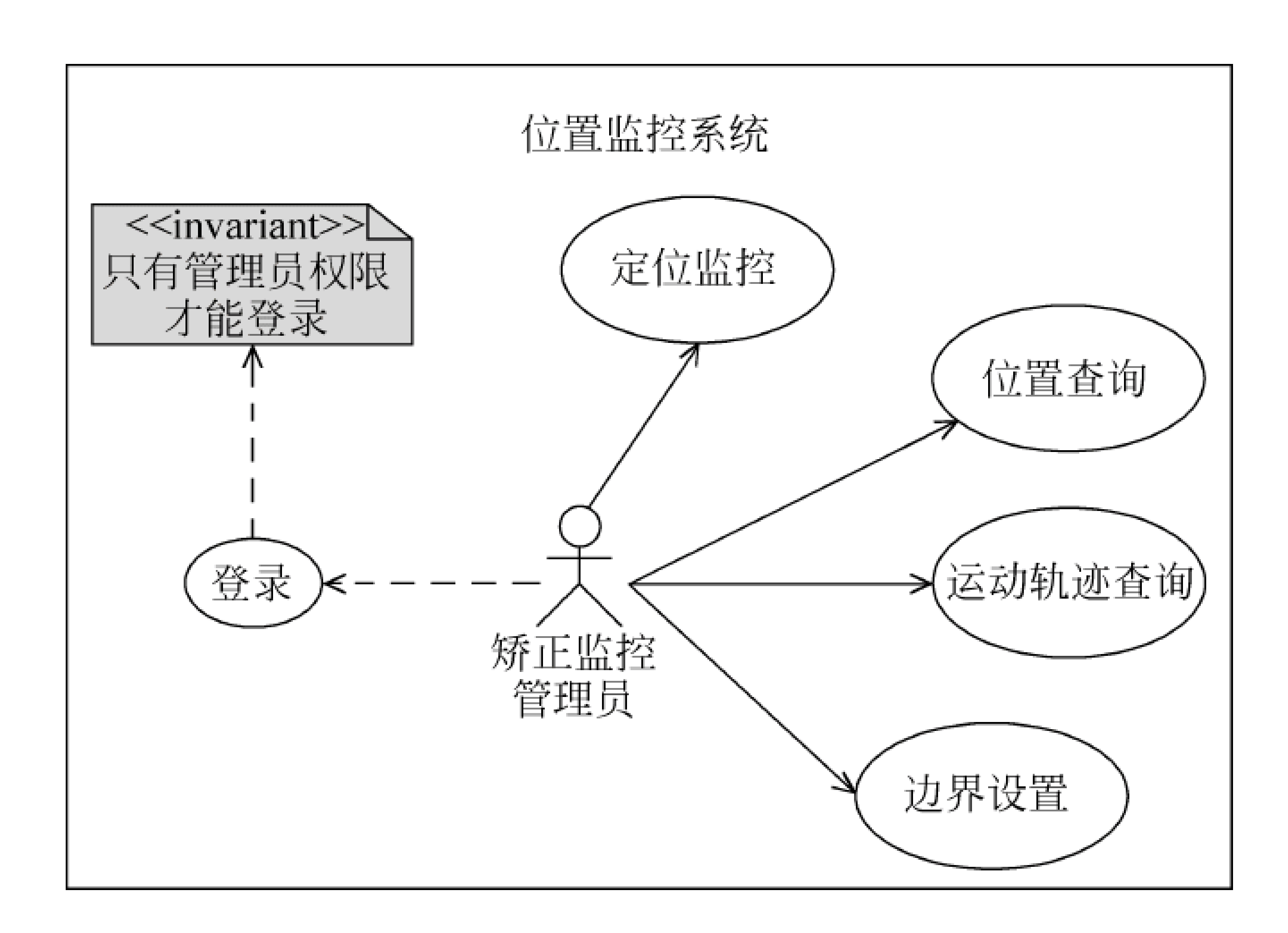


图 6-7 位置监控系统用例模型

图 6-7 中矫正管理员通过身份认证登录系统后的主要操作描述如表 6-2 所示。

表 6-2　位置监控系统用例描述表

用例 ID	JZ-1
用例名称	位置监控系统
参与者	矫正监控管理员
简要说明	通过矫正监控系统，矫正监控管理员能实时查看矫正对象位置信息
前置条件	矫正管理员通过身份认证登录
基本事件流	1. 管理员单击“定位监控”按钮，查看每个矫正对象的位置 2. 管理员单击“位置查询”按钮，输入要查询矫正对象的 ID 或者姓名，系统将搜索该对象长期的位置信息，以列表形式显示出来 3. 管理员单击“运动轨迹查询”按钮，输入要查询的矫正对象的 ID 或姓名以及时间段，系统将显示这段时间的位置变化，并动态显示在地图上 4. 管理员单击“边界设置”按钮，输入矫正对象的 ID 或姓名，然后选择地理范围，最后单击“保存”按钮
其他事件流	边界设置完成若按“取消”按钮，管理员编辑的边界将无效
异常事件流	1. 位置查询以及运动轨迹查询出现异常信息，负责人单击“确认”按钮 2. 返回到初始查询界面
后置条件	定位监控将会显示每个矫正对象在地图上的位置
注释	无

（2）警示告知系统用例模型。

警示告知是通过位置监控来判断并记录矫正人员是否越界以及矫正是否到期等，并对警示进行处理。当矫正对象超出系统设定的安全活动范围，系统会显示报警，并通过手机自动发送报警信息，自动记录警示信息，通知矫正管理员采取一些措施。警示告知系统用例模型如图 6-8 所示。

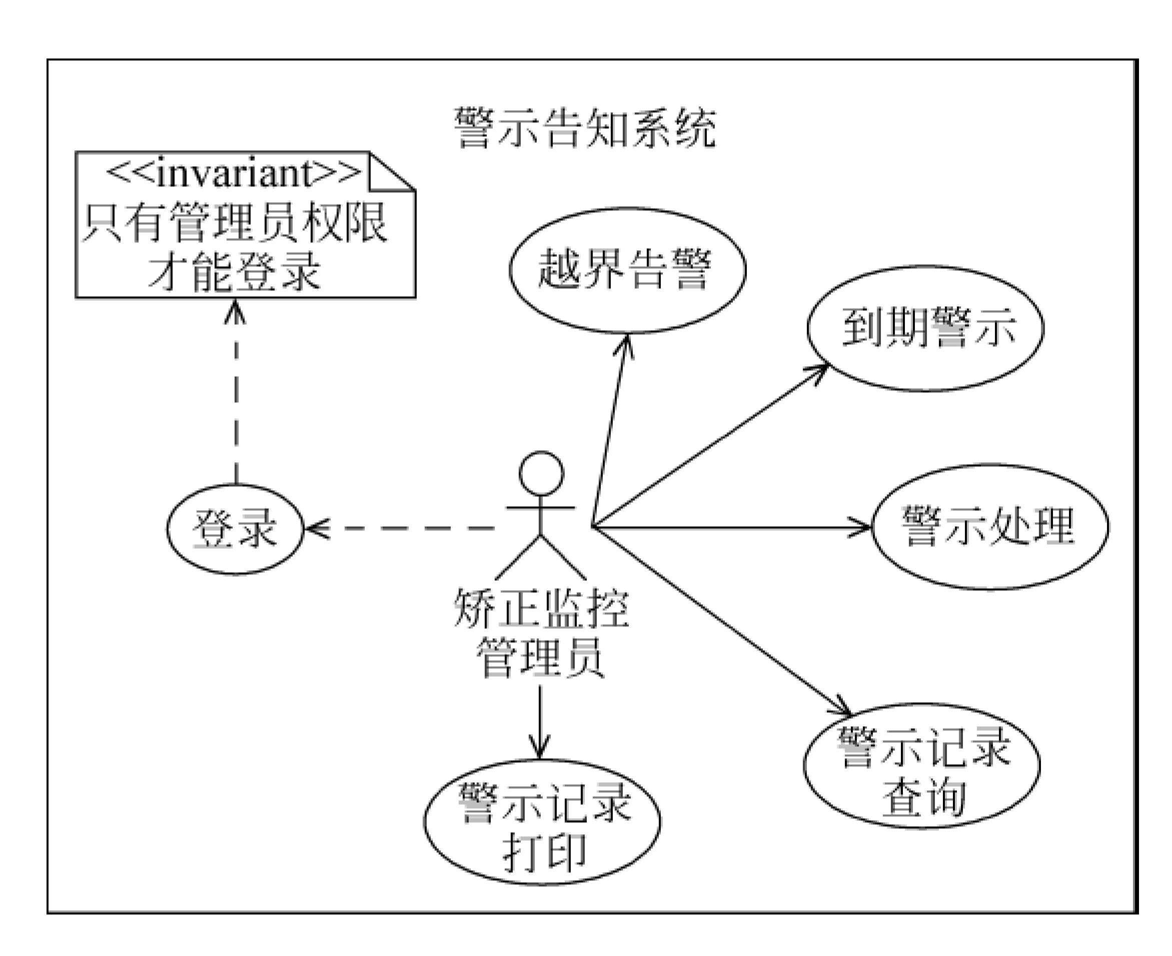


图 6-8　警示告知系统用例模型

图 6-8 中矫正监控管理员通过身份认证登录系统后的主要操作描述如表 6-3 所示。

表 6-3 警示告知系统用例描述表

用例 ID	JZ-2
用例名称	警示告知系统
参与者	矫正监控管理员
简要说明	通过警示告知系统,矫正监控管理员查看矫正对象的警示信息
前置条件	矫正管理员通过身份认证登录
基本事件流	1. 管理员单击“越界告警”按钮,查看报警情况 2. 管理员单击“到期预警”按钮,查看矫正到期人员,并下发解除矫正通知 3. 管理员单击“警示记录查询”按钮,查看已处理警示记录,方便备案 4. 管理员单击“警示记录打印”按钮,将某个人的警示记录打印出来装入档案
其他事件流	矫正监控管理员对于警报信息要及时作出处理,并将处理结果提交到系统,系统将显示在警示处理界面
异常事件流	1. 警示记录查询中出现异常报错,负责人单击“确认”按钮 2. 返回到警示记录查询界面
后置条件	警示处理将显示警示处理的具体细节
注释	无

(3) 档案管理系统用例模型。

档案管理主要用于管理矫正对象的档案信息以及矫正管理员的档案信息,提供基本的操作功能。档案管理同时提供与信息管理系统连接,减少了工作人员的任务量。档案管理系统用例模型如图 6-9 所示。

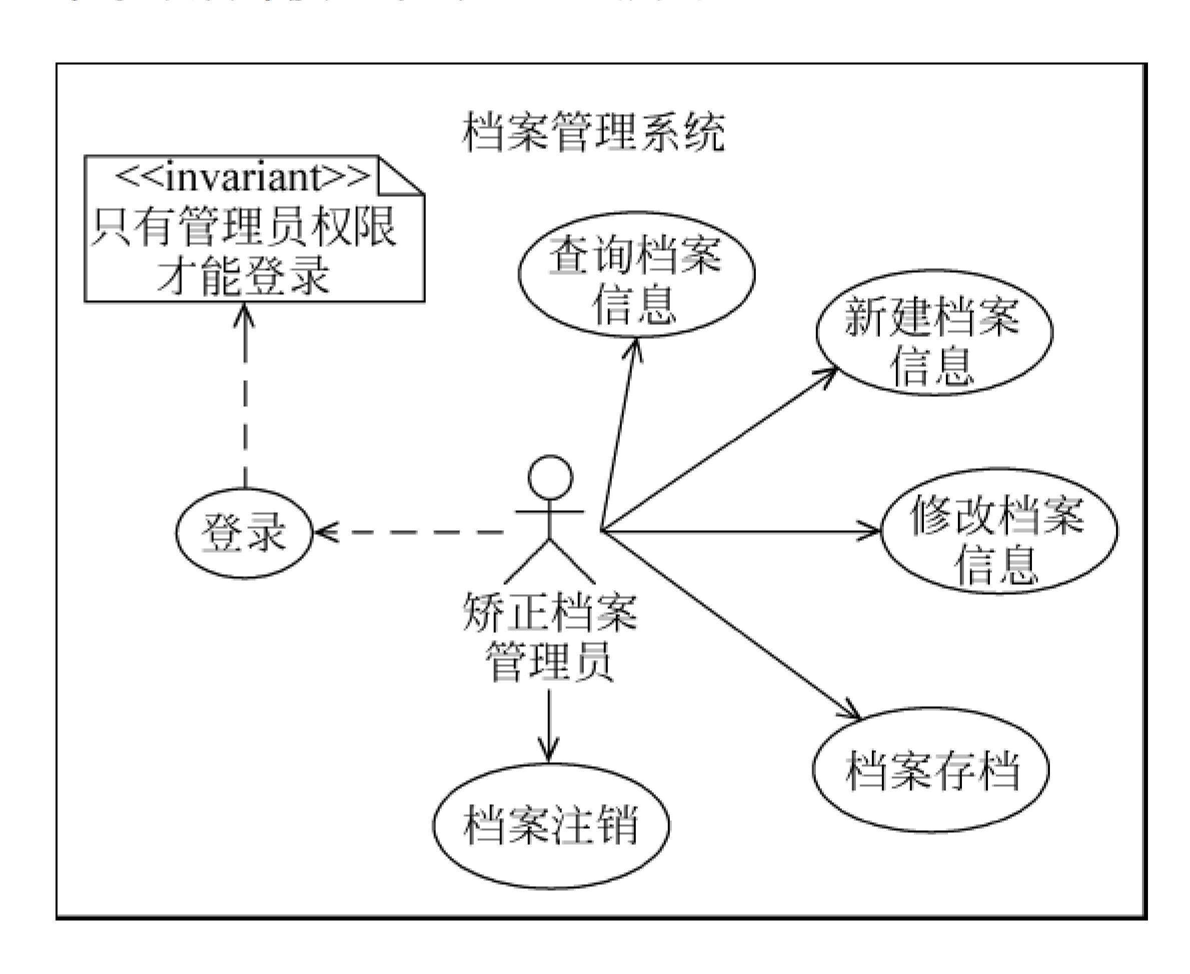


图 6-9 档案管理系统用例模型

表 6-4 是档案管理系统用例描述表，详细描述了用例的功能需求。

表 6-4　档案管理系统用例描述表

用例 ID	JZ-3
用例名称	档案管理系统
执行者	档案管理员
简要说明	建立矫正对象的基本信息和管理员的基本信息，包括其详细的个人资料和家庭信息等
涉众利益	涉众希望每个人员的档案有唯一的 ID
前置条件	管理员已经登录档案管理系统
基本事件流	1. 管理员请求管理矫正对象档案信息 2. 系统显示矫正对象基本信息表 3. 管理员可以进行如下操作： 3.1　查询矫正对象档案 3.2　新建矫正对象档案 3.3　修改矫正对象档案 3.4　矫正对象档案存档 3.5　矫正对象档案注销 4. 管理员填写验证信息并提交 5. 系统验证提交的信息 6. 系统返回操作成功的信息
其他时间流	5. 验证失败，返回到第 4 步
异常事件流	5. 验证管理员权限是否合格；3.2 如果管理员新建档案信息，则自动转入第 2 步
后置条件	系统已经保存了矫正人员的基本信息和管理员的基本信息
注释	3. 档案包括社区矫正文书规则建立

(4) 流程管理系统用例模型。

流程管理系统用来实现矫正管理系统中的业务流程规范，矫正中的各项事务都按照该流程执行，同时系统还具有流程编辑功能，可以查看流程状态，与文书管理系统连接，进行文书的转发和审批。流程管理系统用例模型如图 6-10 所示。

图 6-10 中矫正流程管理员通过身份认证登录系统后的主要操作描述如表 6-5 所示。

表 6-5　流程管理系统用例描述表

用例 ID	JZ-4
用例名称	流程管理系统
参与者	矫正流程管理员
简要说明	通过流程管理系统，矫正流程管理员能对流程规范以及流程进行管理

续表

用例 ID	JZ-4
前置条件	矫正流程管理员通过身份认证登录
基本事件流	1. 管理员单击“流程规范查询”按钮,查看矫正流程规范 2. 管理员单击“流程规范编辑”按钮,即可编辑流程规范 3. 管理员单击“流程审核”按钮,对事务进行审核,并填写审核结果 4. 管理员单击“流程查看”按钮,查看某个事件流程的进展情况,对某个流程还可单击“流程编辑”按钮进行编辑,可以重新定义流程,或者更改某个流程
其他事件流	流程审核是领导对于事务的审核功能,设计不同级别的权限,审核结束要提交审核才能通过 在单击“提交”按钮之前,负责人随时可以单击“返回”按钮,流程编辑所修改的内容都不会影响原来的内容
异常事件流	1. 提示错误信息,负责人单击“确认”按钮 2. 返回到流程浏览界面
后置条件	系统将显示每个事件进行的流程以及进度
注释	无

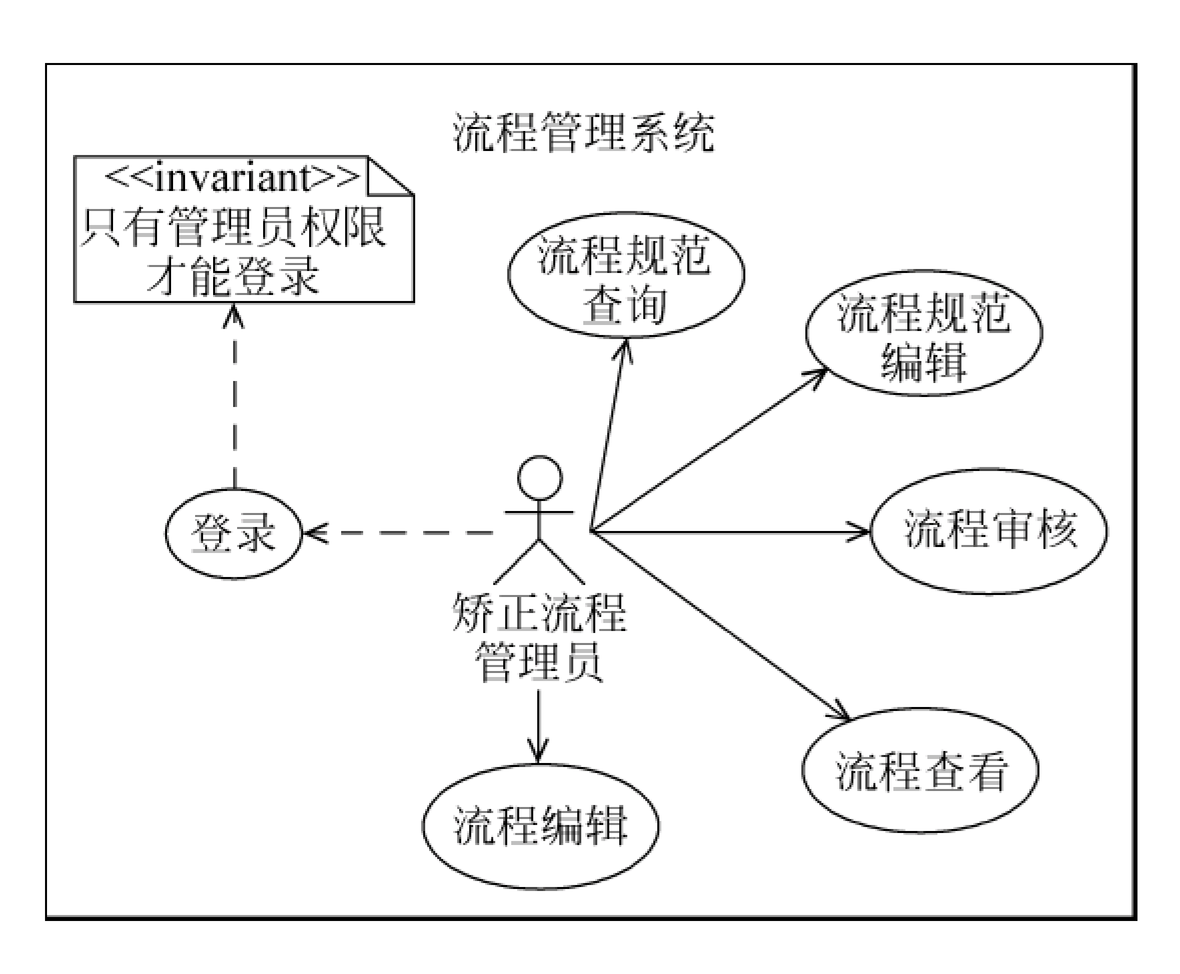


图 6-10　流程管理系统用例模型

(5) 文书管理系统用例模型。

文书管理系统负责文书的管理工作,同时根据管理机构的不同级别,按照流程规范进行文书的传输和流转,而且可以查询文书状态,随时追踪文书的审批时间和人员。文书管理系统用例模型如图 6-11 所示。

图 6-11 中矫正文书管理员通过身份认证后登录系统后的主要操作描述如表 6-6 所示。

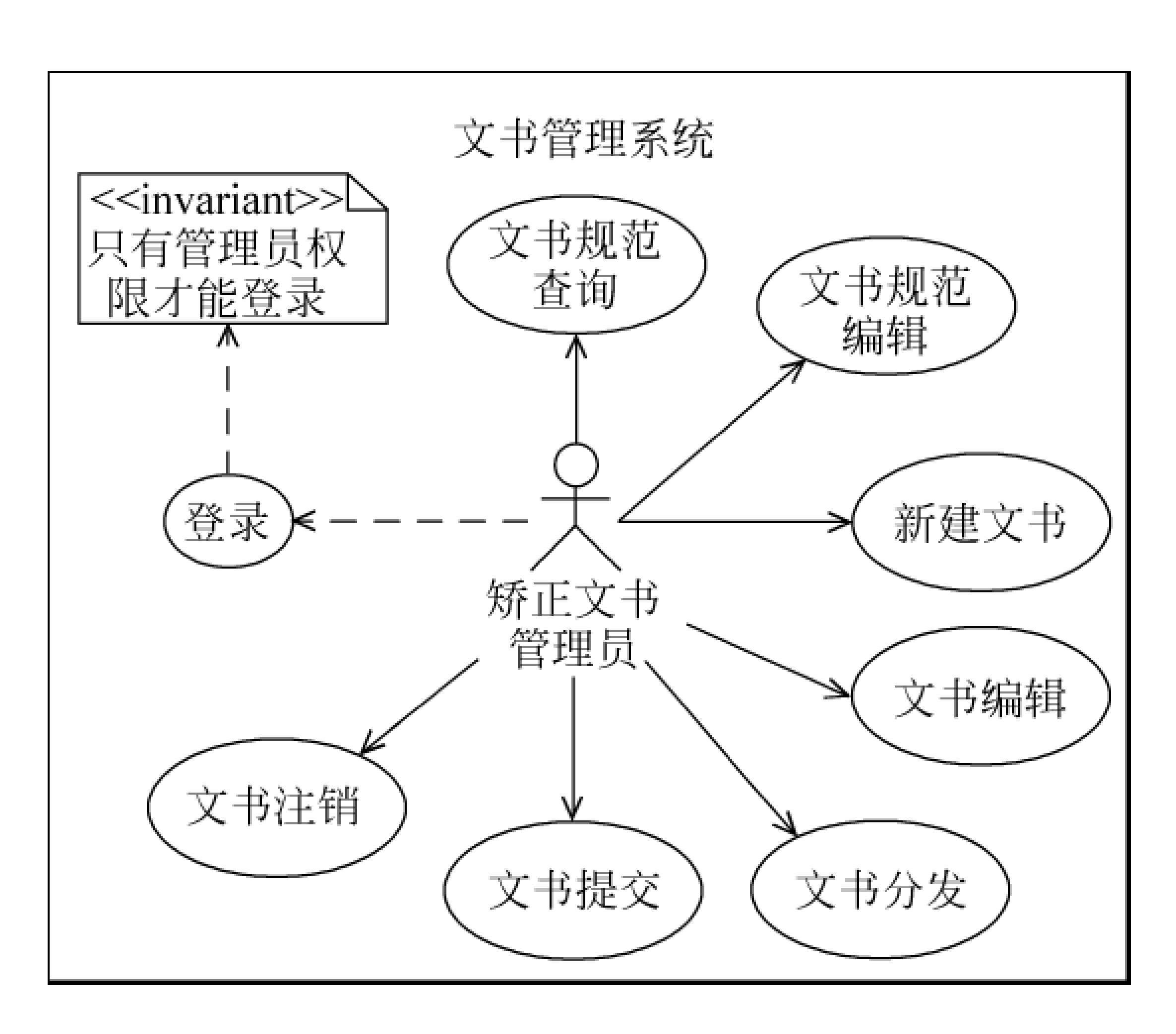


图 6-11 文书管理系统用例模型

表 6-6 文书管理系统用例描述表

用例 ID	JZ-5
用例名称	文书管理系统
参与者	矫正文书管理员
简要说明	通过文书管理系统,矫正文书管理员能管理矫正中的各种文书资料
前置条件	矫正文书管理员通过身份认证登录
基本事件流	1. 管理员单击“文书规范查询”按钮,查看各个文书的规范和要求 2. 管理员单击“文书规范编辑”按钮,可在线编辑某个文书的规范和格式 3. 管理员单击“新建文书”按钮,选择文书模板建立文书 4. 管理员单击“文书编辑”按钮,编辑已建立的文书 5. 管理员单击“文书分发”按钮,将文书向所选部门分发,方便文书传递 6. 管理员单击“文书提交”按钮,审批文书并提交 7. 管理员对作废的文书进行注销
其他事件流	管理员编辑完成后若单击“取消”按钮,将取消编辑
异常事件流	1. 文书提交若出现异常信息,管理员单击“确认”按钮 2. 显示文书提交失败,返回到文书管理首页
后置条件	系统首页显示每个人的矫正文书信息列表
注释	无

(6)日常管理系统用例模型。

日常管理系统包括集中学习、公益劳动、矫正情况汇报以及矫正考核等内容,主要负责矫正对象素质的提高和培养。日常管理系统用例模型如图 6-12 所示。

图 6-12 中矫正日常管理员通过身份认证登录系统后的主要操作描述如表 6-7 所示。

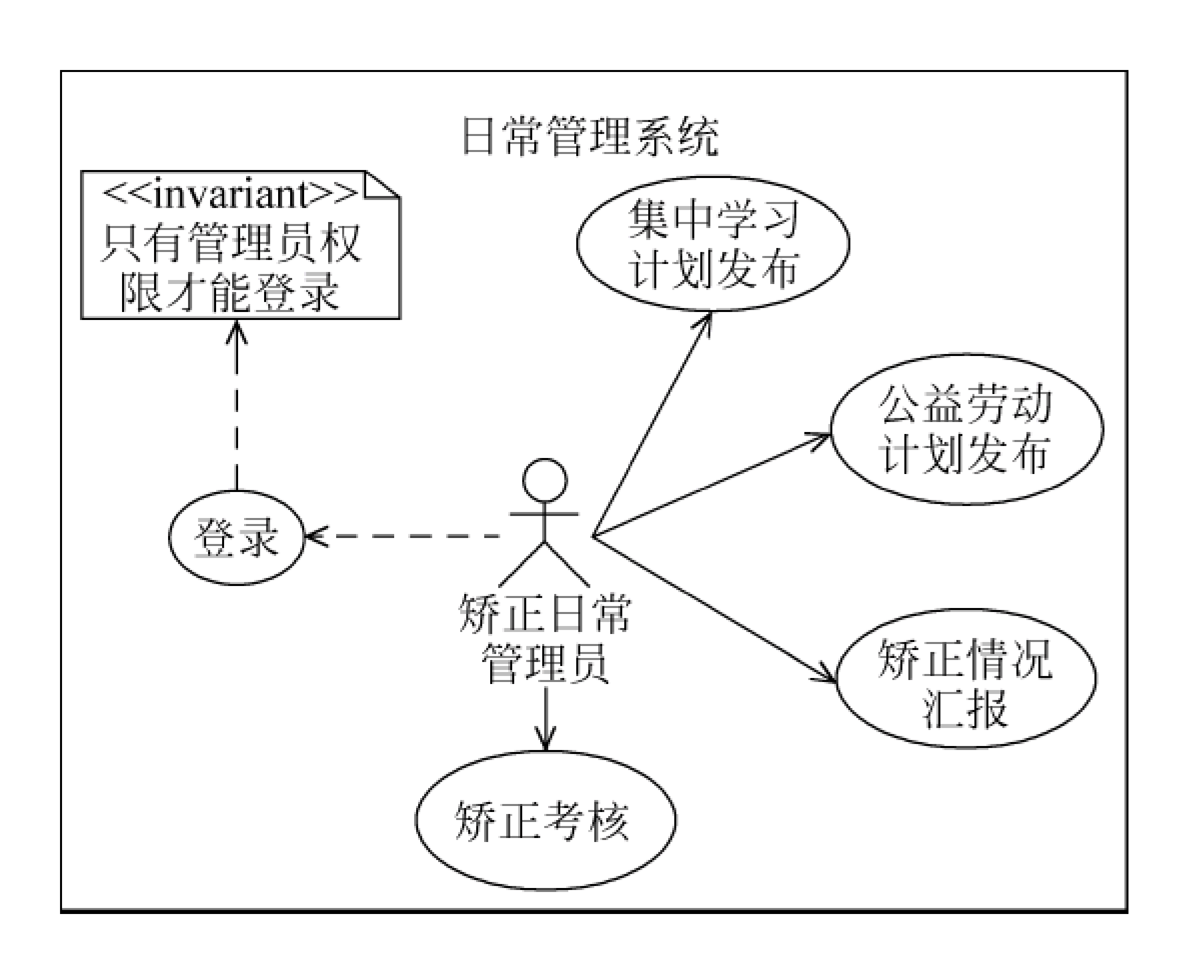


图 6-12 日常管理系统用例模型

表 6-7 日常管理系统用例描述表

用例 ID	JZ-6
用例名称	日常管理系统
参与者	矫正日常管理员
简要说明	通过日常管理系统，矫正日常管理员制定矫正计划以及考核矫正情况
前置条件	矫正日常管理员通过身份认证登录
基本事件流	1. 管理员单击“集中学习计划发布”按钮，可以编辑并发布学习计划 2. 管理员单击“公益劳动计划发布”按钮，编辑并发布公益劳动计划 3. 管理员单击“矫正情况汇报”按钮，定时对矫正对象出现的各项情况进行汇报 4. 管理员单击“矫正考核”按钮，可新建、查看矫正对象的考核情况
其他事件流	1 和 2 计划发布前单击“保存”按钮即可保存在草稿箱，等待发布 1 和 2 计划发布前若单击“取消”按钮，可取消计划发布
异常事件流	1. 系统若出现异常错误信息，管理员单击“确认”按钮 2. 返回系统首页，不会对系统信息进行修改
后置条件	管理员通过日常管理系统考核矫正对象日常工作
注释	无

（7）权限管理用例模型。

权限管理包括权限查询、权限修改、增加权限功能，由矫正小组组长修改矫正管理员的各项权限。权限管理系统用例模型如图 6-13 所示。

图 6-13 中矫正小组组长通过身份认证登录系统后的主要操作描述如表 6-8 所示。

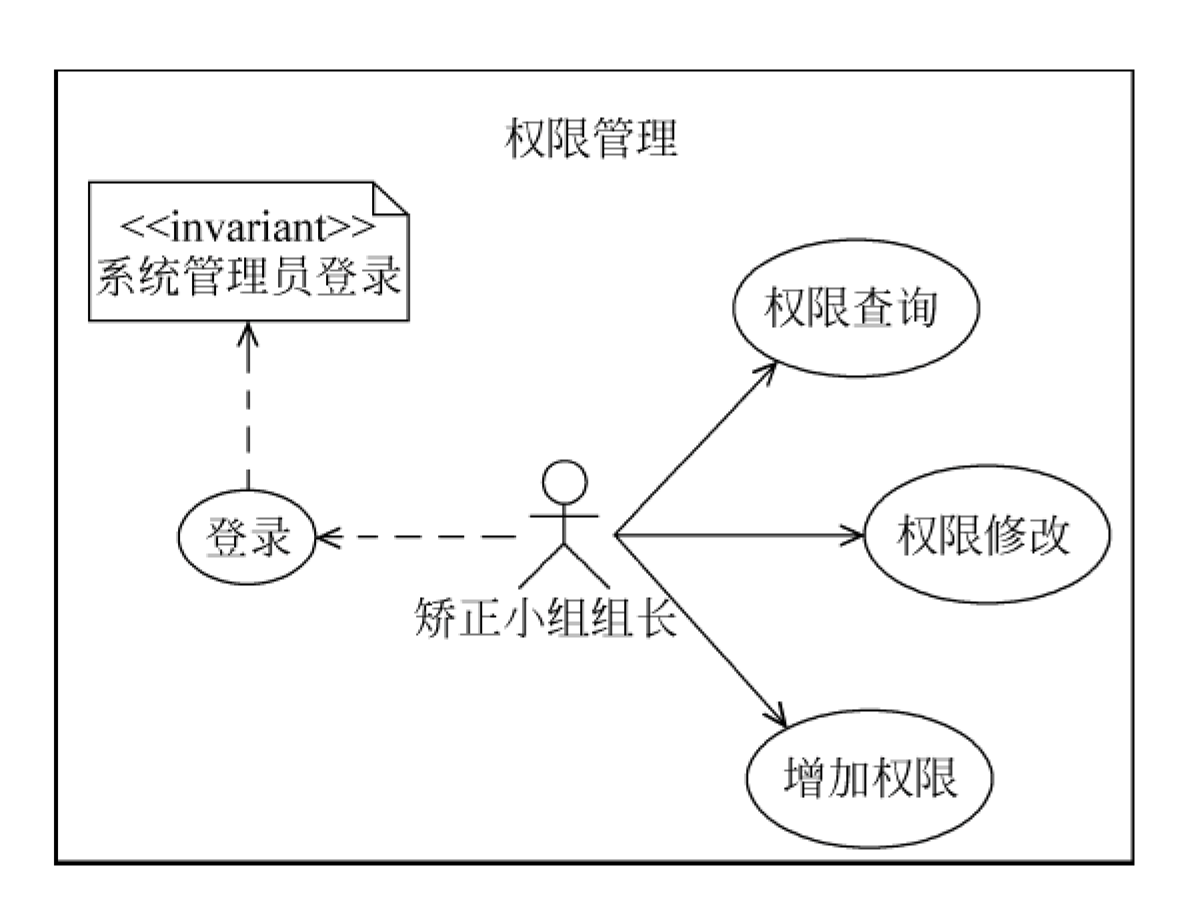


图 6-13　权限管理系统用例模型

表 6-8　权限管理用例描述表

用例 ID	JZ-7
用例名称	权限管理
参与者	矫正小组组长
简要说明	矫正小组通过权限管理来管理各个用户的权限
前置条件	矫正小组组长通过身份认证登录
基本事件流	1. 组长单击“权限查询”按钮,查看每个用户的权限 2. 组长选择某个用户,单击“权限修改”按钮,然后修改其权限 3. 组长单击“增加权限”按钮,可添加某个用户的权限
其他事件流	操作 2 中若单击“取消”按钮,则取消对某个用户权限的修改
异常事件流	1. 若系统出现异常错误信息,组长单击“确认”按钮 2. 返回系统首页,对系统信息不会影响
后置条件	修改完权限将返回用户的权限列表
注释	无

(8)工作机制用例模型。

工作机制是本系统与现有政府职能系统之间链接机制,可方便用户进行访问。工作机制用例模型如图 6-14 所示。

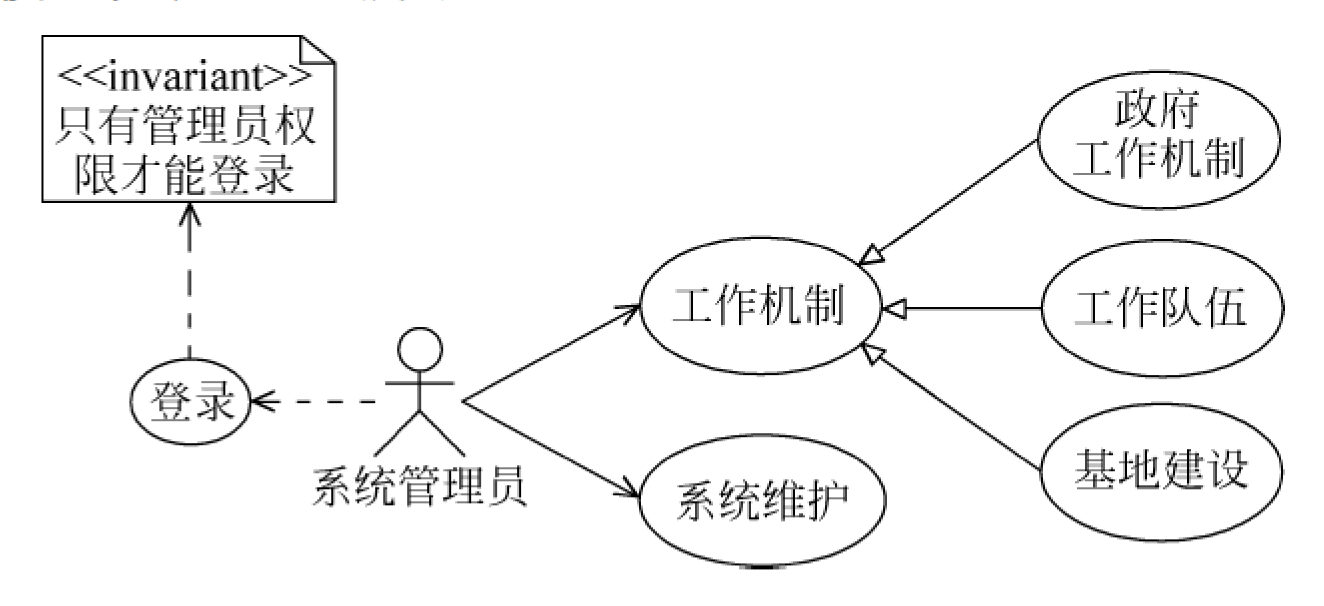


图 6-14　工作机制用例模型

图 6-14 中系统管理员通过身份认证登录系统后的主要操作描述如表 6-9 所示。

表 6-9　工作机制用例描述表

用例 ID	JZ-8
用例名称	工作机制
参与者	系统管理员
简要说明	通过工作机制，系统管理员可添加系统接口，与政府系统之间进行连接
前置条件	系统管理员通过身份认证登录
基本事件流	1. 系统管理员通过后台添加接口，与政府、司法、户政、公安等系统进行对接 2. 系统管理员对系统进行错误处理等维护
其他事件流	无
异常事件流	无
后置条件	系统的工作机制显示外部系统链接
注释	无

6.3.2　系统设计

社区矫正管理信息系统包含大量的矫正信息，包括矫正对象的位置信息、档案信息、邮件信息、文书信息以及流程等，其数据量大、信息类型多样化。使用 UML 对社区矫正管理信息系统进行面向对象的分析和设计，可以从系统的底层把握矫正管理过程中的特征，为下一步系统开发打好基础。在社区矫正管理信息系统建立模型时要设计处理大量的模型元素，如类、接口、组件、节点、图等，将语义上相近的模型元素组织在一起，就构成了包，包是从较高层次来组织的一种信息系统模型。

系统主要由以下八个包组成。

(1) 用户接口包(com. personnel. jsp)：此包位于其他包的顶层，为系统用户提供可视化界面和操作访问接口与服务。系统基于 Java 开发，用户接口包采用 jsp 技术，同时搭建 ExtJS 作为显示层的框架。

(2) 控制模型包(com. personnel. actions)：该包包含了创建业务控制器的 Action 类，用来处理用户提交的请求，包含矫正监控类 JzjcAction、信息管理类 MessageAction、用户类 UserAction、警示告知类 WarninfoAction 以及个人简历类 GrjlAction 等。此包也是 Struts2 框架的基本组成单元，与 struts. xml 中 action 域的 class 属性对应。

(3) 过程域模型包(com. personnel. service)：此包对系统模块功能进行封装，其中包括位置信息服务 LocationinfoService、矫正监控服务 JzjcService、信息服务

MessageService、用户管理服务 UserService、警示信息服务 WarninfoService、个人简历服务 GrjlService 以及机构管理服务 OrganizationService 等类。

(4) 过程域模型实现包(com. personnel. service. impl)：包含实现 com. personnel. service 包中类的方法。

(5) 抽象 DAO 模型包(com. personnel. dao)：把系统类的方法封装成接口，实现一个表在一个项目中的全部操作。包括位置信息接口 LocationinfoDao、矫正监控接口 JzjcDao、信息接口 MessageDao、用户管理接口 UserDao、警示信息接口 WarninfoDao、个人简历接口 GrjlDao 以及机构管理接口 OrganizationDao 等。

(6) 抽象 DAO 模型实现包(com. personnel. dao. impl)：定义 DAO 接口的具体类，扩展 HibernateDaoSupport，实现 dao 包中类的方法。

(7) 域模型包(com. personnel. model)：将数据库字段封装成对象，包含每个字段的 set 和 get 方法以及基于 Hibernate 映射的配置文件。

(8) 通用工具包(com. personnel. common)：包含了系统用到的通用功能类，如页数统计及分页 PageBean 类、PageCtr 类、页面跳转 SendPage 类、中文转英文类 ChineseCharToEn 类等。

系统各个包之间的调用关系如图 6-15 所示。

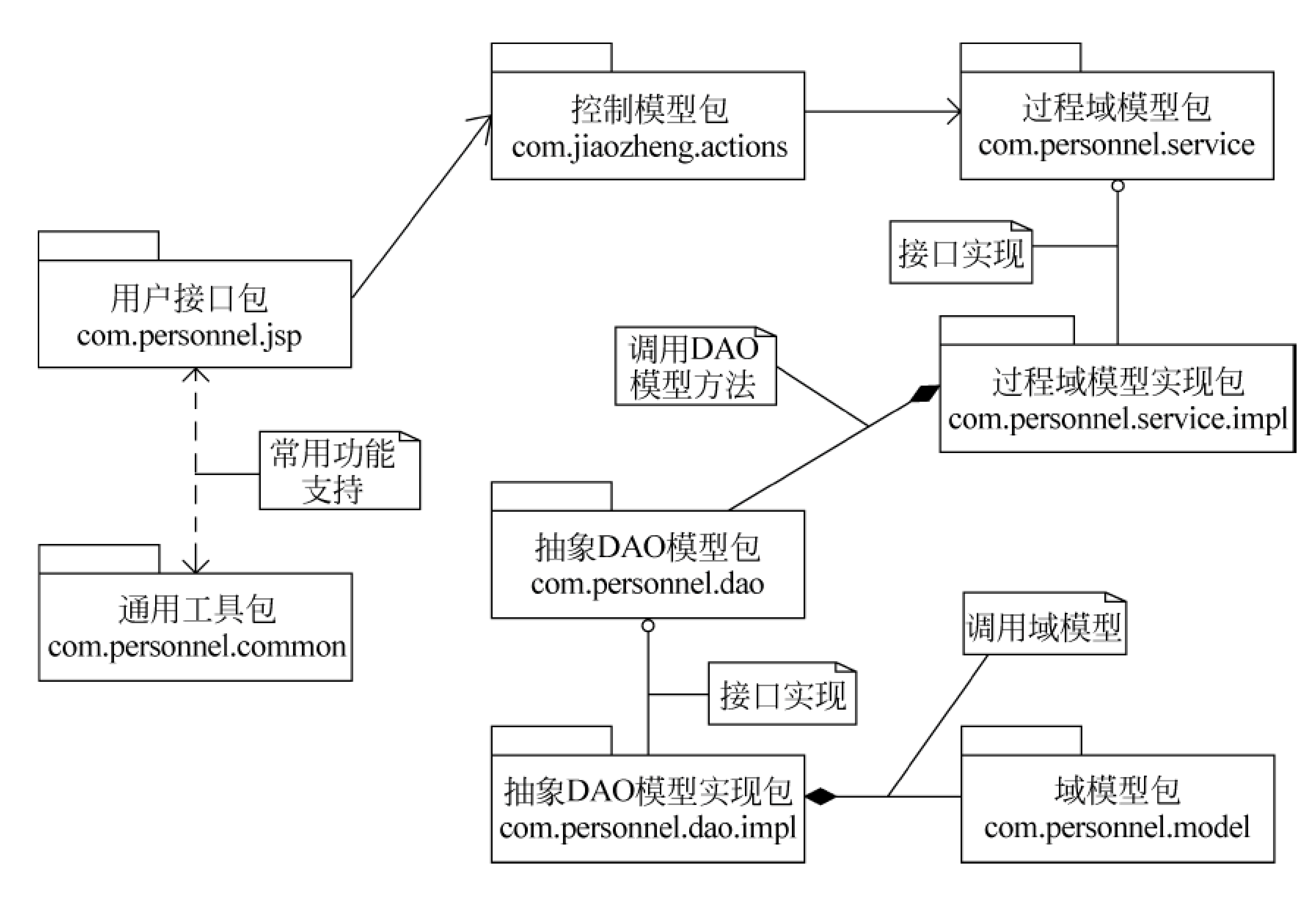


图 6-15　系统包调用关系图

1. 系统主要工作流程分析

《社区矫正实施办法》规定了社区矫正过程中必须遵循的一些业务流程，包含各项事务处理的业务流程标准，本节介绍社区矫正过程中几个主要的流程。

1）调查评估流程

（1）业务流程。

调查评估流程图如图 6-16 所示。

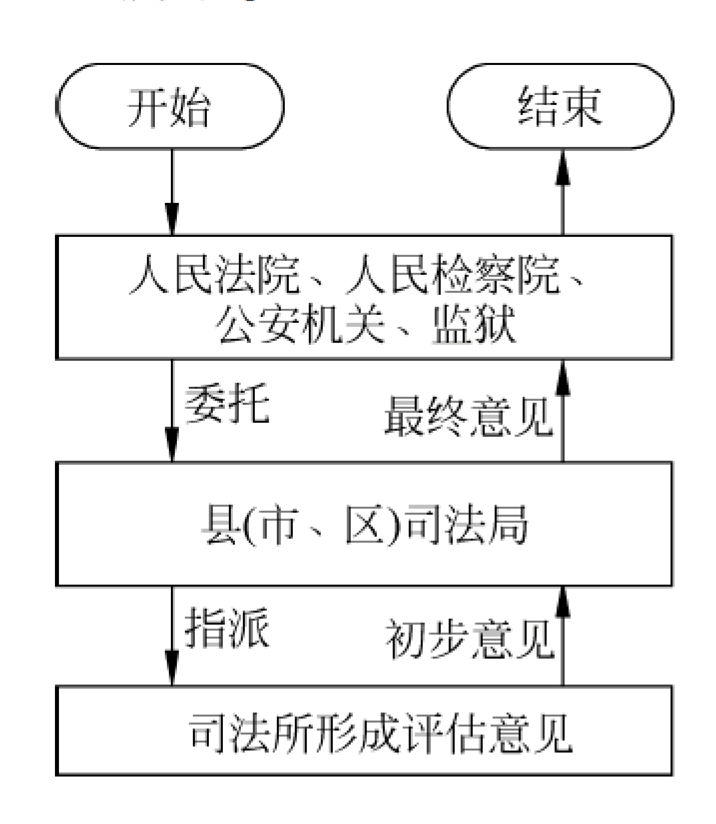


图 6-16　调查评估流程图

（2）流程说明。

① 人民法院、人民检察院、公安机关、监狱对拟适用社区矫正的被告人、罪犯，需要调查其对所居住社区有影响的，可以委托县级司法行政机关进行调查评估。

② 受委托的司法行政机关应当根据委托机关的要求，对被告人或者罪犯的居所情况、家庭和社会关系、一贯表现、犯罪行为的后果和影响、居住地村（居）民委员会和被害人意见、拟禁止的事项等进行调查了解，形成评估意见，及时提交至委托机关。

2）矫正衔接流程

（1）业务流程。

矫正衔接流程的业务流程图如图 6-17 所示。

（2）流程说明。

① 系统通过和法院（预留）、公安机关（预留）、监狱部门的业务系统对接，实现社区矫正人员信息的自动交换。

② 对于适用社区矫正的罪犯，人民法院、公安机关、监狱应当核实其居住地，在向其宣判时或者在其离开监所之前，书面告知其到居住地县级司法行政机关报到的时间期限以及逾期报到的后果，并通知居住地县级司法行政机关；在判决、裁定生效起三个工作日内，送达判决书、裁定书、决定书、执行通知书、假释证明书副本等法律文书，同时抄送其居住地县级人民检察院和公安机关。

③ 对于法院、公安机关、监狱发送纸质材料的情况，系统在县（市、区）司法局提供对衔接的纸质材料进行接收确认和登记功能，登记后向下转派。

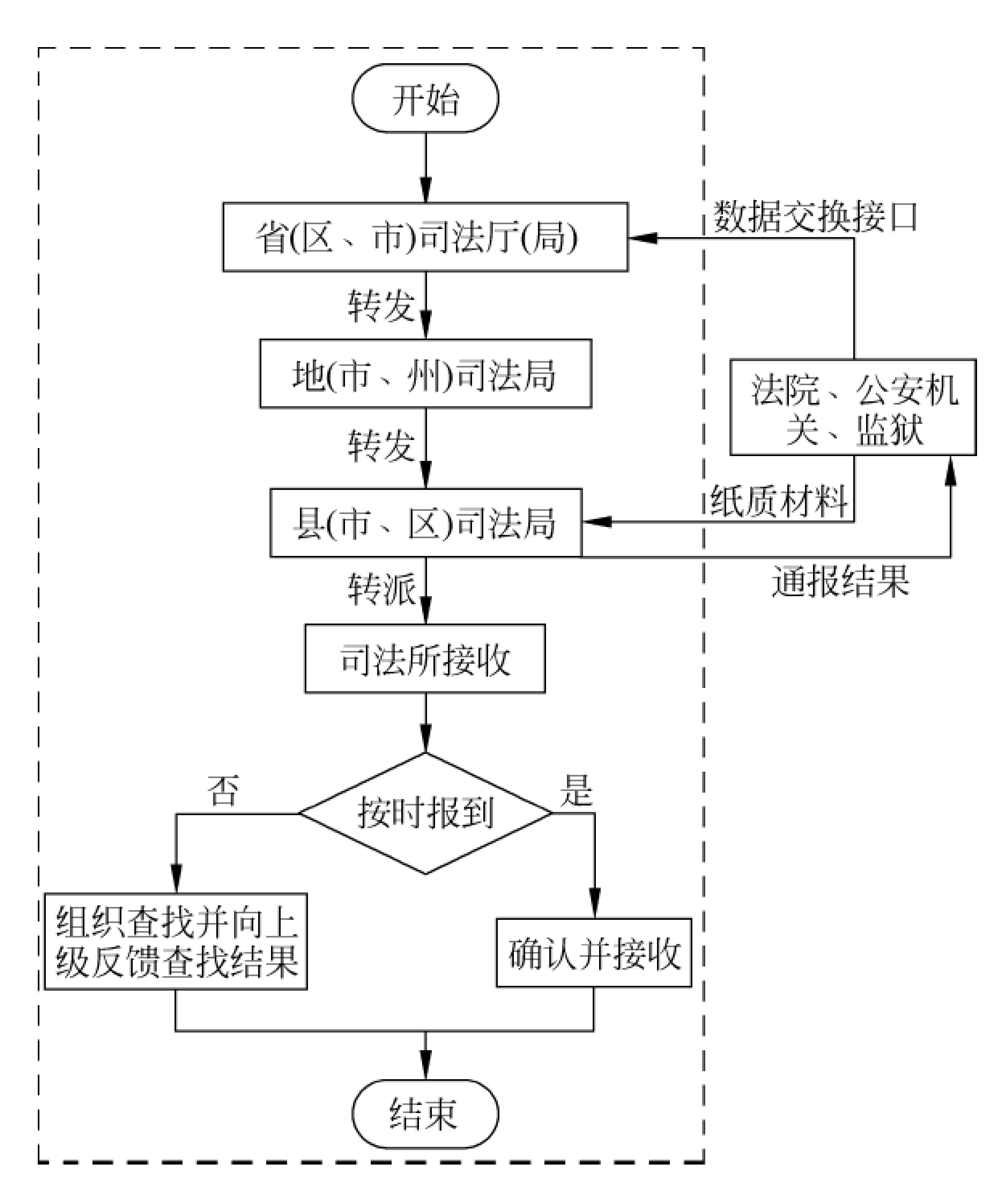


图 6-17　矫正衔接流程图

④ 县级司法行政机关收到法律文书后，应当在三个工作日内送达回执。

⑤ 发现社区矫正人员未按规定时间报到的，县级司法行政机关应当及时组织查找，并通报决定机关。

3）外出审批流程

（1）业务流程。

外出审批流程的业务流程图如图 6-18 所示。

（2）流程说明。

① 社区矫正人员未经批准不得离开所居住的市、县（旗）。

② 社区矫正人员因就医、家庭重大变故等原因，确需离开所居住的市、县（旗），在七日以内的，应当报经司法所批准；超过七日的，应当由司法所签署意见后报经县级司法行政机关批准。

③ 返回居住地时，应当立即向司法所报告。

④ 社区矫正人员离开所居住市、县（旗）不得超过一个月。

4）居住地变更审批流程

（1）业务流程。

居住地变更的业务流程图如图 6-19 所示。

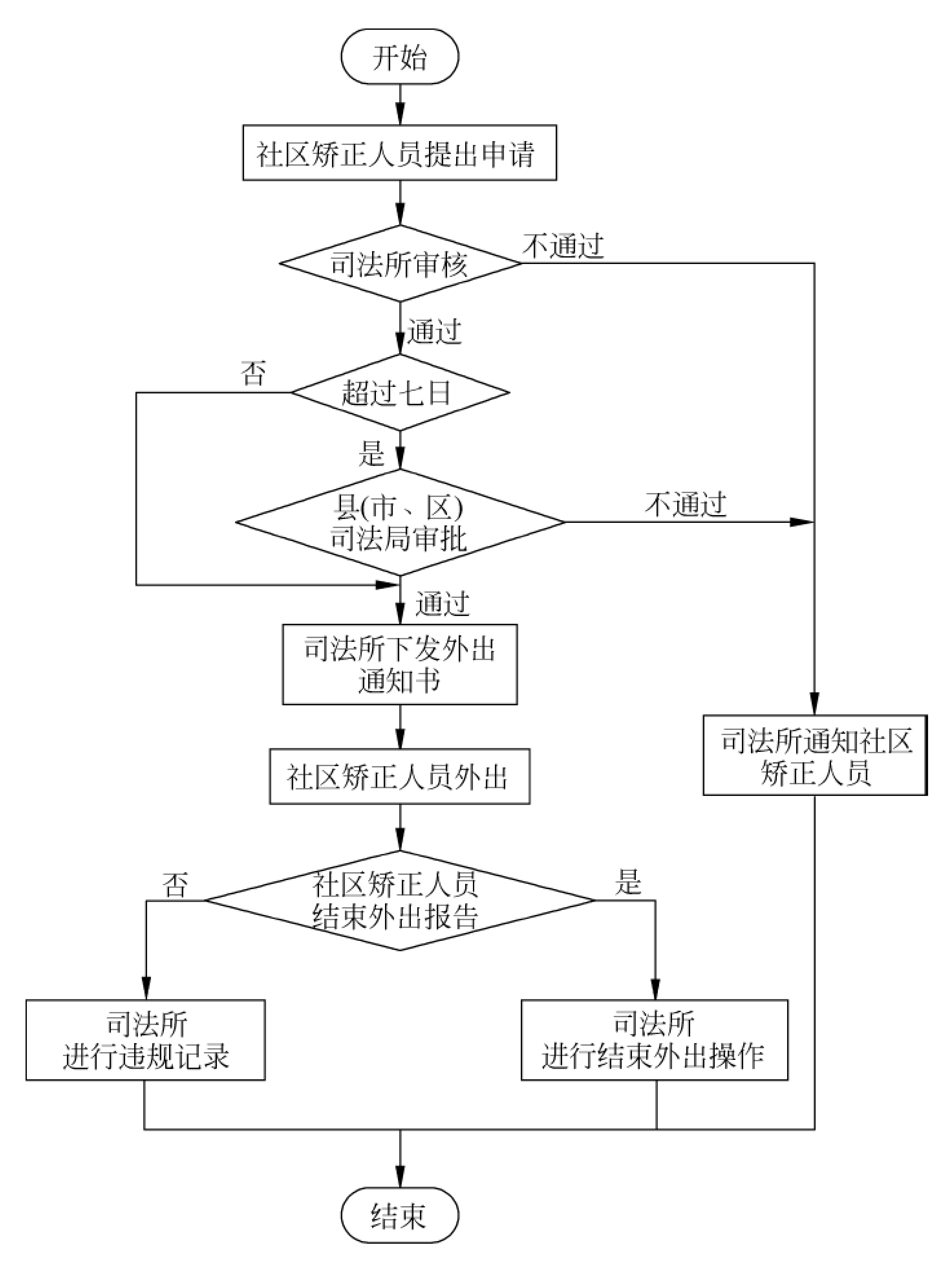


图 6-18 外出审批流程图

（2）流程说明。

① 社区矫正人员未经批准不得变更居住的县（市、区、旗）。

② 社区矫正人员因居所变化确需变更居住地的，应当提前一个月提出书面申请，由司法所签署意见后报经县级司法行政机关审批。

③ 县级司法行政机关在征求社区矫正人员新居住地县级司法行政机关的意见后作出决定。

④ 经批准变更居住地的，县级司法行政机关应当自做出决定之日起三个工作日内，将有关法律文书和矫正档案移交新居住地县级司法行政机关。

⑤ 有关法律文书应当抄送现居住地及新居住地县级人民检察院和公安机关。

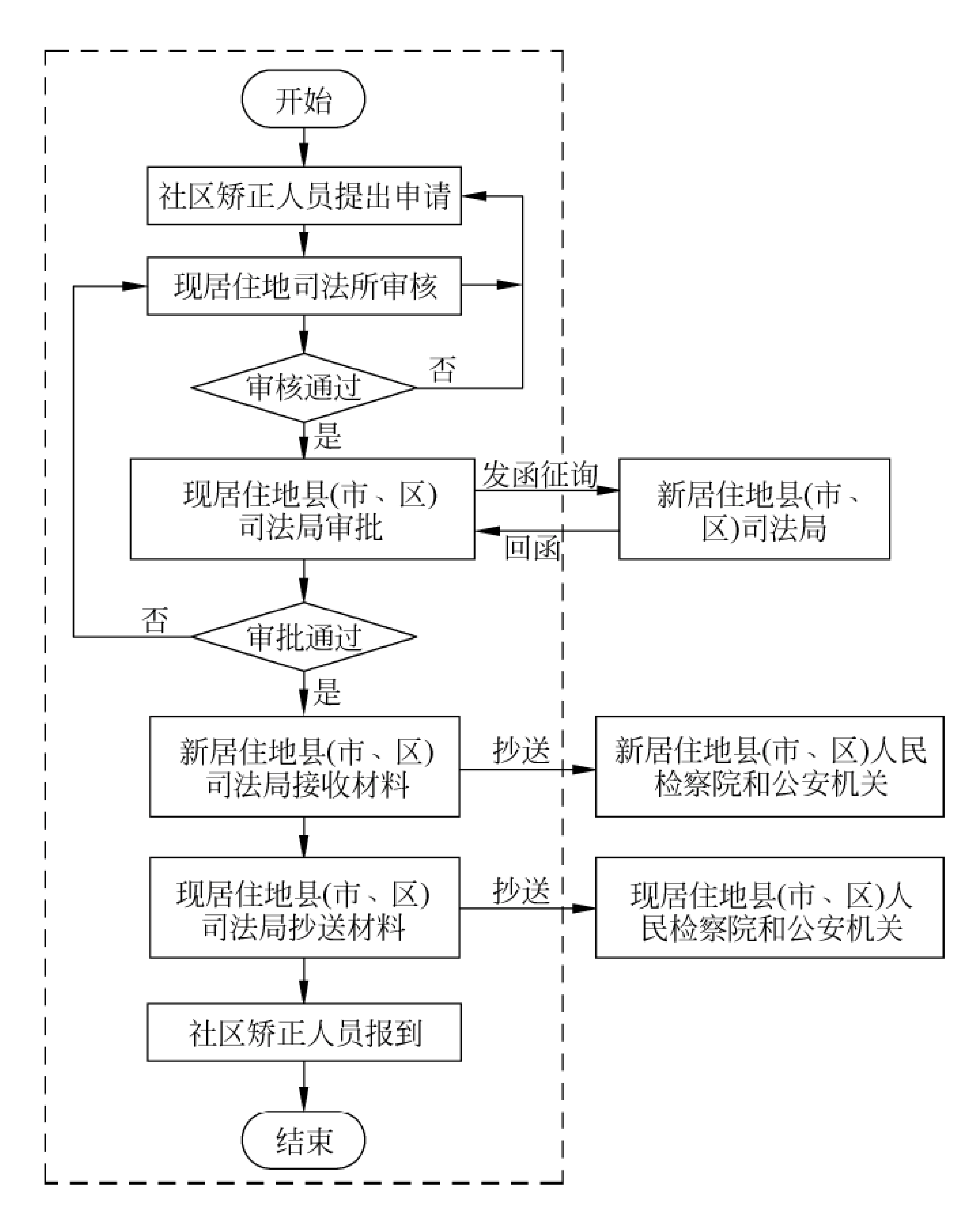


图 6-19　居住地变更流程图

⑥ 社区矫正人员应当自收到决定之日起七日内到新居住地县级司法行政机关报到。

5）警告审批流程

（1）业务流程。

警告审批的业务流程图如图 6-20 所示。

（2）流程说明。

社区矫正人员有下列情形之一的，县级司法行政机关应当给予警告，并出具书面决定：

① 未按规定时间报到的。

② 违反关于报告、会客、外出、居住地变更规定的。

③ 不按规定参加教育学习、社区服务等活动，经教育仍不改正的。

④ 保外就医的社区矫正人员无正当理由不按时提交病情复查情况，或者未经批准进行就医以外的社会活动且经教育仍不改正的。

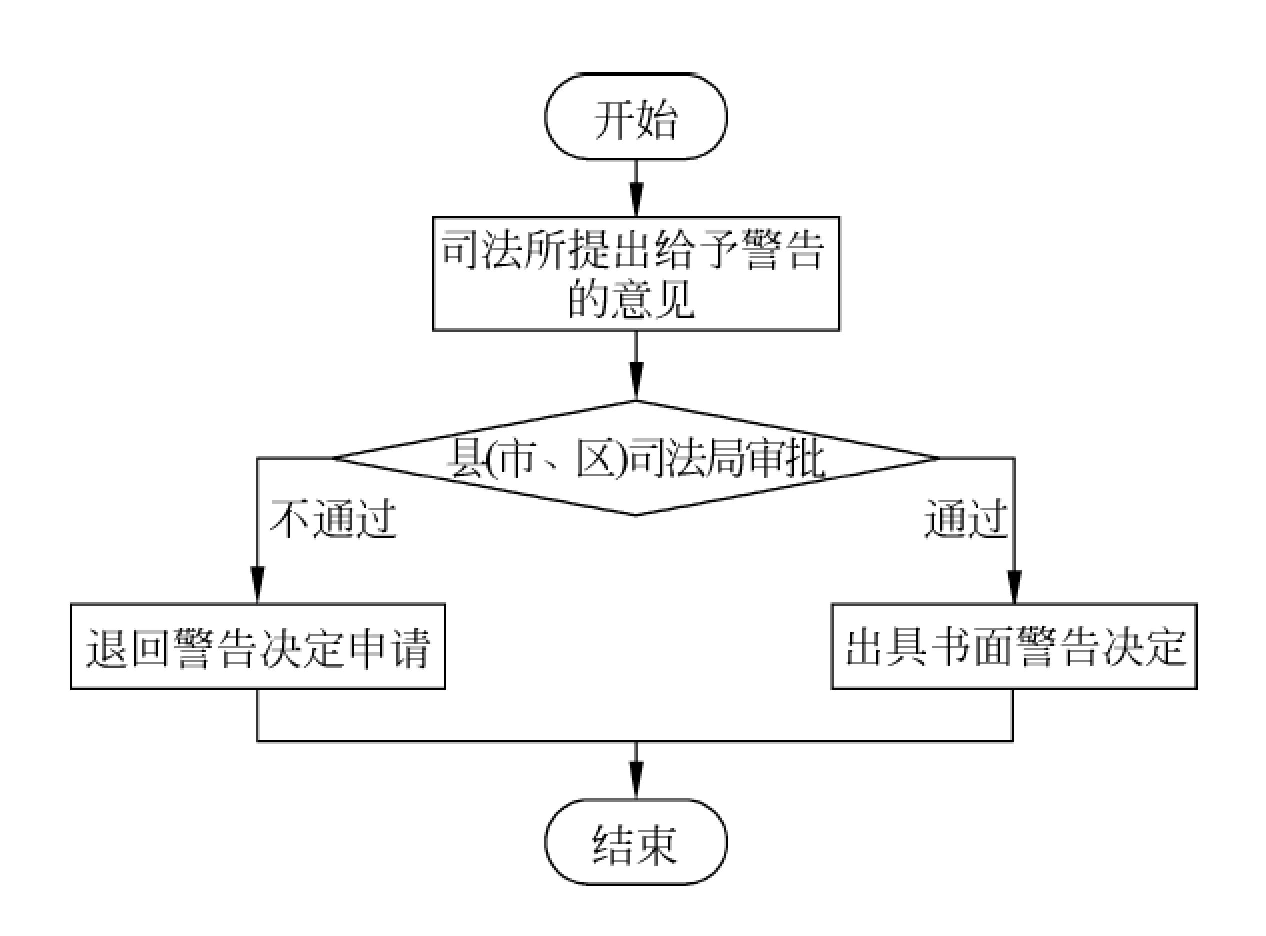


图 6-20　警告审批流程图

⑤ 违反人民法院禁止令，情节轻微的。

⑥ 其他违反监督管理规定的。

6）解除矫正流程

（1）业务流程。

① 社区矫正人员期满解矫流程，如图 6-21 所示。

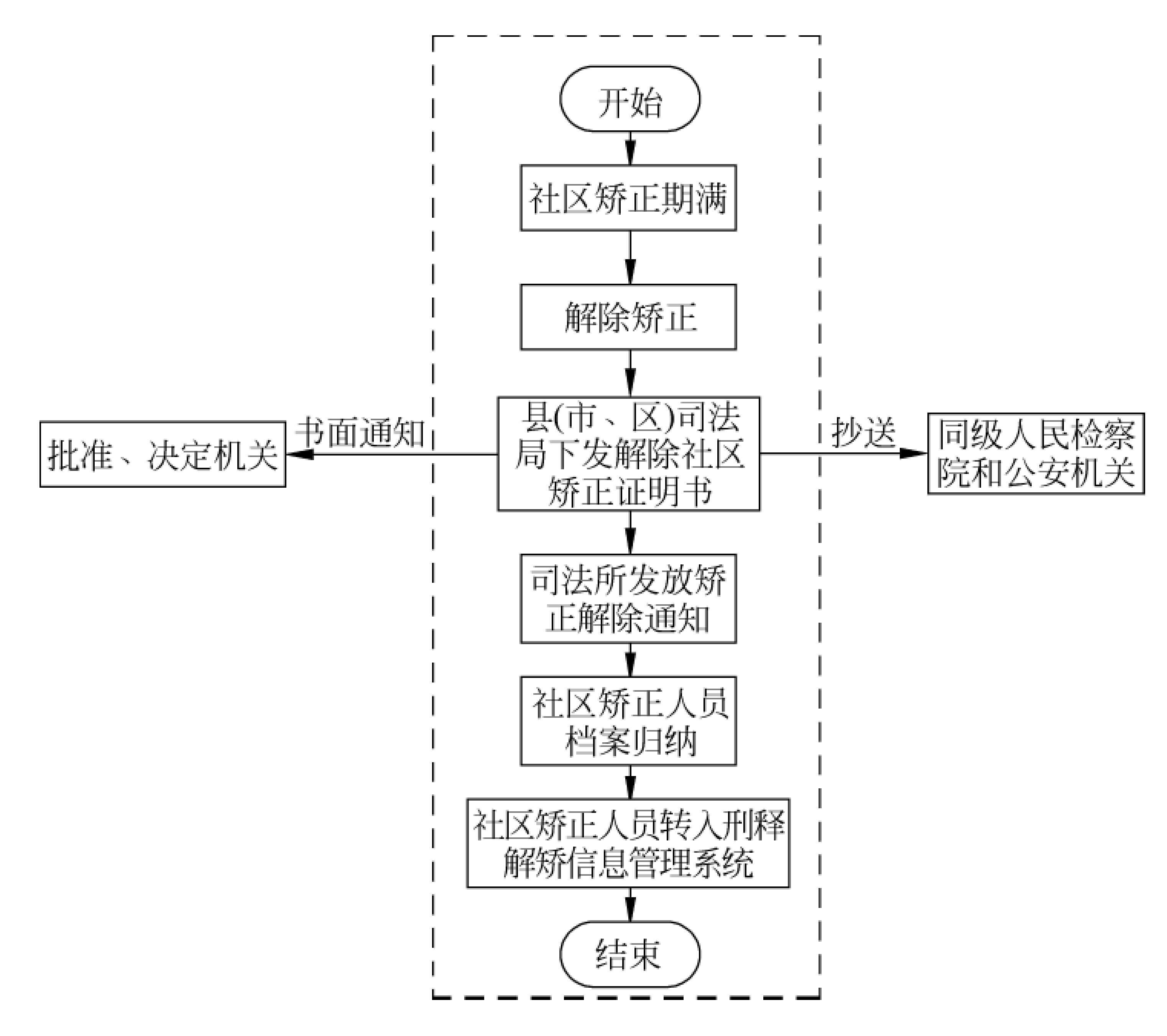


图 6-21　期满解矫流程图

② 暂予监外执行社区矫正人员期满解矫流程，如图 6-22 所示。

图 6-22 暂于监外执行期满解矫流程图

(2) 流程说明。

① 社区矫正人员矫正期满，司法所应当组织解除社区矫正宣告。宣告由司法所工作人员主持，按照规定程序公开进行。

② 县级司法行政机关应当向社区矫正人员发放解除社区矫正证明书，并书面通知决定机关，同时抄送县级人民检察院和公安机关。

③ 暂予监外执行的社区矫正人员刑期届满的，由监狱、看守所依法为其办理刑满释放手续。

7) 终止矫正流程

(1) 业务流程。

终止矫正的业务流程如图 6-23 所示。

(2) 流程说明。

① 社区矫正人员死亡、被决定收监执行或者被判处监禁刑罚的，社区矫正终止。社区矫正人员在社区矫正期间死亡的，县级司法行政机关应当及时书面通知

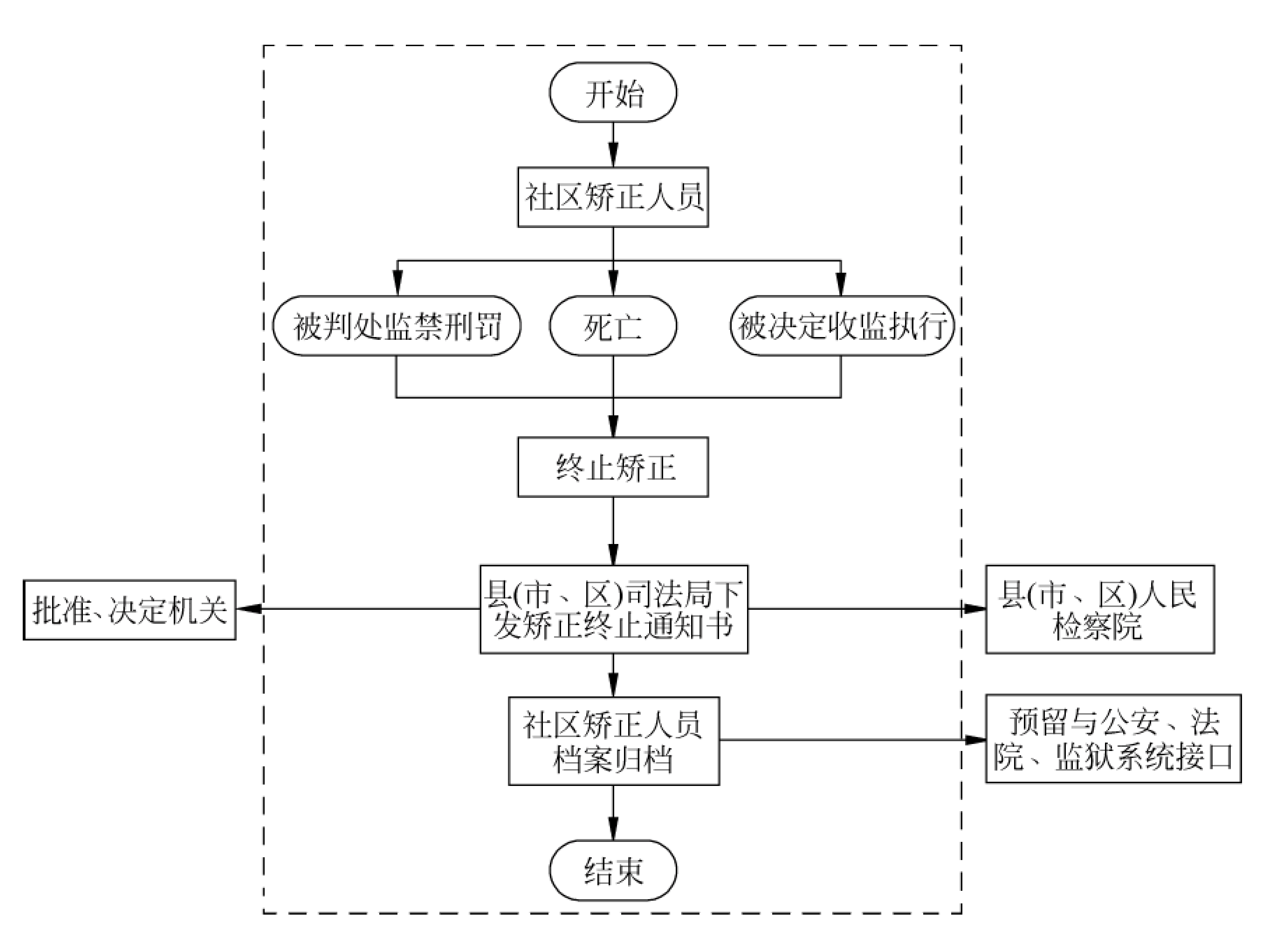


图 6-23 终止矫正流程图

批准、决定机关，并通报县级人民检察院。

② 通过数据交换接口，实现与公安、法院、监狱等业务系统的数据交换，最终实现对收监、羁押、发现漏罪等情况的自动终止社区矫正。

③ 对于社区矫正人员死亡的，司法所工作人员录入相关证明，终止矫正，档案自动归档，不可更改。

2. 社区矫正管理信息系统数据库设计

本系统主要用来管理矫正对象的档案信息、管理员信息、位置信息、矫正信息、通信信息以及机构信息等，这些信息是严格按照中华人民共和国司法行业标准《社区矫正管理信息系统技术规范》规定的数据采集结构规范要求设计的，并在此基础上进行了扩展。本系统除了系统的一些数据表以及机构表之外，矫正数据库有二十九个表，包含了各类数据信息。这里介绍一些主要的数据表。

1）档案管理数据表

社区矫正人员基本信息采集表（info）：包括矫正人员编号、管理对象类别、姓名、曾用名、性别、民族证件类型、证件号码、出生日期等各项详细信息共有 98 个字段，其中某些字段有相应的条文和参考格式标准。该数据表是矫正对象档案管理中重要的数据，如表 6-10 和表 6-11 所示。

表 6-10　社区矫正人员基本信息采集表(Info)

数据元名称	标　　识	类型	是否空	相关条文和标准
社区矫正人员编号	SQJZRYBH	C20	否	编码参照 SF 05001—2013
管理对象类别	GLDXLB	C2	否	编码参照 SF 05001—2013
姓名	XM	C20	否	自填
曾用名	CYM	C20	是	自填
性别	XB	C2	否	参照国标 GB/T 2261.1—2003
民族	MZ	C20	否	参照国标 GB/T 3304—1991
证件类型	ZJLX	C20	否	自填
证件号码	ZJHM	C18	否	自填
出生日期	CSRQ	D	否	自填,格式“YYYY-MM-DD”
有无护照	YWHZ	TN	否	自填
护照号码	HZHM	C20	是	自填
有无回乡证	YWHXZ	TN	否	自填
回乡证号码	HXZHM	C20	是	自填
有无台胞证	YWTBZ	TN	否	自填
台胞证号码	TBZHM	C20	是	自填
文化程度	WHCD	C20	否	编码参照 SF 05001—2013
健康状况	JKZK	TN	否	编码参照 SF 05001—2013
具体健康状况	JTJKZK	C100	是	自填
是否有传染病史	SFYCRBS	TN	否	自填
具体传染病史	JTCRBS	C100	是	自填
心理是否健康	XLSFJK	TN	否	自填
…	…	…	…	…
矫正小组人员组成情况	JZXZRYZCQK	TN	否	编码参照 SF 05001—2013
是否使用定位管理	SFSYDWGL	TN	否	自填
终端类型	ZDLX	TN	是	自填
终端号码	ZDHM	C20	是	自填

表 6-11　矫正小组成员信息采集表

数据元名称	标　　识	类型	是否空	相关条文和标准
社区矫正人员编号	SQJZRYBH	C20	否	编码参照 SF 05001—2013
小组成员类型	XZCYLX	C20	否	编码参照 SF 05001—2013
小组成员类别	XZCYLB	DT	否	编码参照 SF 05001—2013
姓名	XM	C20	否	自填
性别	XB	C2	否	参照国标 GB/T 2261.1—2003
出生日期	CSRQ	D	否	自填,格式“YYYY-MM-DD”
学历	Xl	C2	是	参照国标 GB/T 4658—2006
最高学位	ZGXW	C3	是	参照国标 GB/T 6864—2003

续表

数据元名称	标　　识	类型	是否空	相关条文和标准
政治面貌	ZZMM	C2	是	参照国标 GB/T 4762—1984
专业	ZY	C7	是	参照国标 GB/T 13745—2009
职业	ZY	C20	是	参照国标 GB/T 6565—1999
工作单位	GZDW	C100	是	自填
联系电话	LXDH	C20	是	自填，格式为固话“区号-号码”
手机	SJ	C11	否	自填
家庭住址	JTZZ	C50	是	自填
社会工作专业类职称	SHGZZYLZC	C2	是	编码参照 SF 05001—2013
薪酬水平段	XCSPD	C2	是	编码参照 SF 05001—2013
合同期	HTQ	C2	是	编码参照 SF 05001—2013
备注	BZ	C500	是	自填

2）定位监控数据表

位置监控是社区矫正过程中一项非常重要的环节，获取定位监控数据也至关重要。定位数据主要包括矫正人员编号、矫正人员终端编号、经度、纬度、定位状态、定位时间以及服务提供商。这些信息为位置展示提供了数据，通过这些数据能够直观地将地理位置显示在百度地图上供管理员查看，如表 6-12 所示。

表 6-12　社区矫正人员定位历史信息采集表

数据元名称	标　　识	类型	是否空	相关条文和标准
社区矫正人员编号	SQJZRYBH	C50	否	编码参照 SF 05002—2013
社区矫正人员终端号码	SQJZRYZDHM	C50	否	
经度	JD	N8	否	
纬度	WD	N8	否	
定位状态	DWZT	TN	否	编码参照 SF 05002—2013
定位时间	DWSJ	DT8	否	
服务提供商	FWS	TN	否	编码参照 SF 05002—2013

3）信息管理数据表

信息管理可以给矫正对象发送邮件信息、短信信息通知相关事项，邮件基本信息采集表包括信息编号、邮件标题、具体内容、发件人编号、接收人编号、创建时间以及邮件类型等，如表 6-13 所示。

表 6-13　邮件基本信息采集表

数据元名称	标　　识	类　　型	是否允许空
信息编号	Id	Int	否
邮件标题	Title	Varchar	是
具体内容	Content	Text	是
发件人编号	FromUserId	Int	是
接收人编号	recepUserId	Int	是
创建时间	CreateTime	Datetime	是
邮件类型	msgType	Int	是
标记	Sign	Int	是

短信发送基本信息采集表如表 6-14 所示。

表 6-14　短信发送基本信息采集表

数据元名称	标　　识	类　　型	是　否　空	相关条文和标准
发送人员编号	FSRYBH	C50	否	
用户类型	YHLX	C50	否	
接收号码	JSHM	C20	否	
发送内容	FSNR	C400	是	
发送时间	FSSJ	DT2	否	
短信编号	DXBH	C50	否	
服务提供商	FWS	TN	否	编码参照 SF 05002—2013
短信状态	DXZT	C50	否	编码参照 SF 05002—2013

短信发送基本信息采集表包含发送方号码、发送内容、接收时间、短信编号、服务提供商以及短信状态等，如表 6-15 所示。

表 6-15　接收短信基本信息采集表

数据元名称	标　　识	类　　型	是否允许空	相关条文和标准
发送方号码	FSHM	C20	否	
发送内容	FSNR	C400	是	
接收时间	JSSJ	DT8	否	
短信编号	DXBH	C50	否	
服务提供商	FWS	TN	否	编码参照 SF 05002—2013
短信状态	DXZT	C50	否	编码参照 SF 05002—2013

4）矫正管理数据表

矫正管理负责矫正的一些事务管理，包含矫正解除、矫正日常考核、矫正学习、矫正期间的各项事务等。其中以矫正解除为例，矫正解除包含矫正人员编号、矫正

类型、矫正解除日期、收监执行原因、收监执行日期等二十多个字段，如表 6-16 所示。

表 6-16　矫正解除(终止)信息采集表

数据元名称	标识	类型	是否空	相关条文和标准
社区矫正人员编号	SQJZRYBH	C20	否	编码参照 SF 05001—2013
矫正解除(终止)类型	JJLX	C20	否	编码参照 SF 05001—2013
矫正解除(终止)日期	JJRQ	D	否	格式“YYYY-MM-DD”
收监执行原因	SJZXYY	TN	是	编码参照 SF 05001—2013
收监执行类型	SJZXLX	TN	是	编码参照 SF 05001—2013
收监执行日期	SJZXRQ	D	是	自填，格式“YYYY-MM-DD”
死亡日期	SWSJ	D	否	自填，格式“YYYY-MM-DD”
死亡原因	SWYY	C20	否	编码参照 SF 05001—2013
具体死因	SWYY	C500	否	自填
矫正期间表现	JZQJBX	TN	否	编码参照 SF 05001—2013
认罪态度	RZTD	TN	否	编码参照 SF 05001—2013
矫正期间是否参加职业技能培训	SFCJZYJNPX	TN	否	自填
矫正期间是否获得职业技能证书	SFHDZYJNZS	TN	否	自填
技术特长及等级	JSTCJDJ	C500	是	自填
是否三无人员	SFSWRY	TN	否	自填
危险性评估	WXXPG	TN	否	编码参照 SF 05001—2013
家庭联系情况	JTLXQK	TN	否	编码参照 SF 05001—2013
矫正期间特殊情况备注及帮教建议	TSQKBZJBJJY	C500	是	自填
备注	BZ	C500	是	自填
司法所解除人	SFSJCR	C20	是	自填
司法所解除时间	SFSJCSJ	DT	是	自填，格式“YYYY-MM-DD”

6.3.3　系统详细设计及实现

1. 系统技术支持框架集成

社区矫正管理信息系统的技术框架如图 6-24 所示。

各功能组件的功能如下：

1）表示层

表示层采用 Struts2 的 MVC 模式，用户与表示层进行交互的过程是通过 Web 客户端发出请求，该请求被接收后由前端控制器对该请求进行判断和识别，再把请

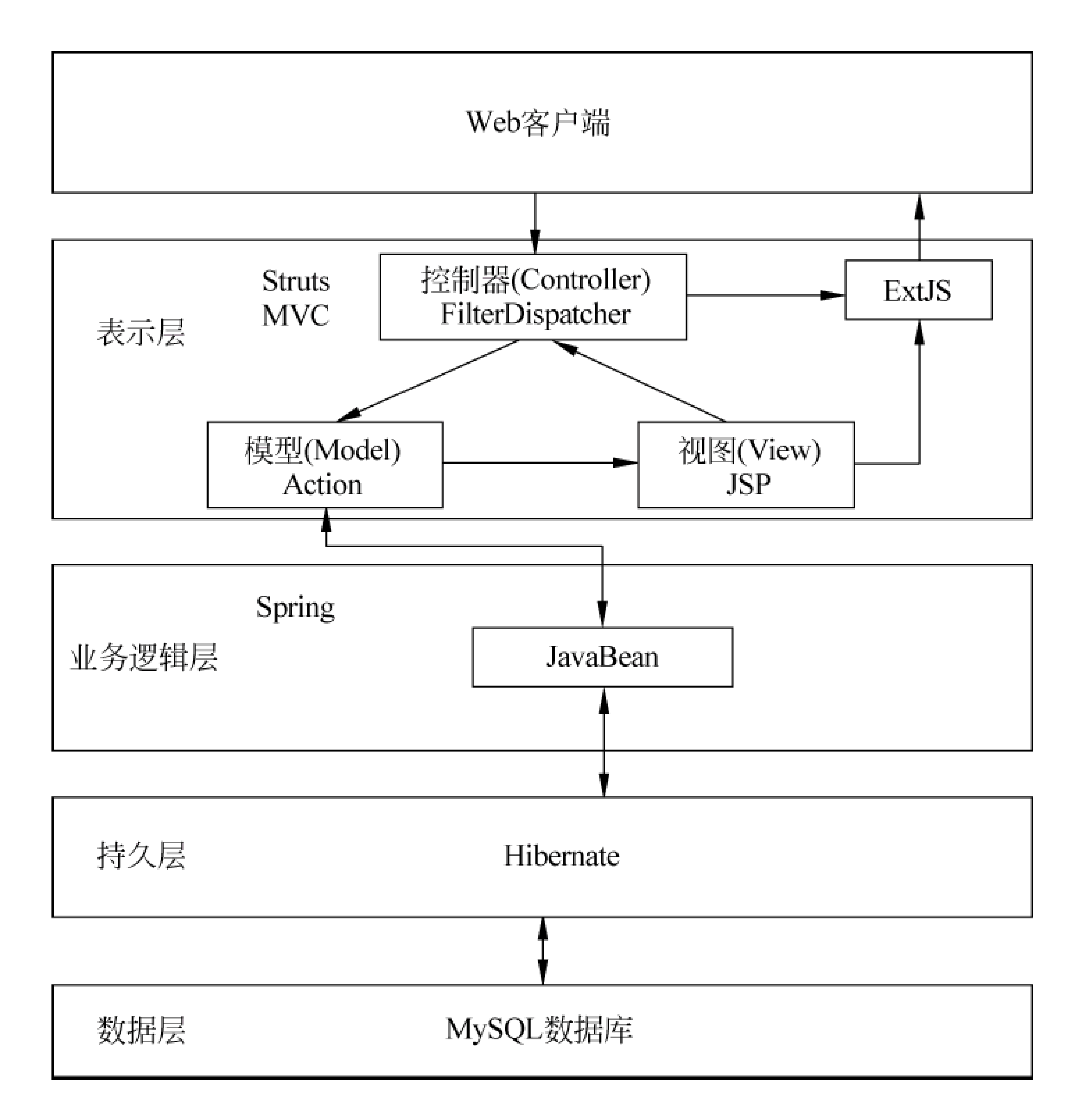


图 6-24　社区矫正管理信息系统技术支持框架图

求分配到与之对应的控制器,控制器调用相应的业务接口,这些业务接口由业务逻辑层实现。

2）业务逻辑层

业务逻辑层采用 Spring 框架,在该层可以方便地组织业务逻辑,同时进行事务管理。在矫正管理信息系统中,使用依赖注入的方式,利用 IOC(Inversion of Control,控制反转)容器,通过配置 Spring 的配置文件 applicationContextx. xml,配置相应的 bean,管理 bean 的对应关系,这样大大降低了业务逻辑层和持久层之间的耦合。

3）持久层

本系统采用 Hibernate 框架作为持久层的 ORM 框架,并使用数据访问对象模式。它能有效降低业务逻辑层和数据层的耦合程度,使我们可以将注意力专注于业务逻辑。

2. 位置监控模块

社区矫正过程中实现位置监控的关键是定位技术。尽管社区矫正位置监控系统采用的定位技术有基于 GPS、基于移动运营商基站、基于北斗等,但应用开发只

需要借助于GPS定位实现其定位功能。来源于移动运营商的社区矫正人员信息资源库仅包含人员编号、终端号等信息。社区矫正位置监控系统则具有二维地图基本操作、定位信息获取处理、网络通信等功能，如图6-25所示。

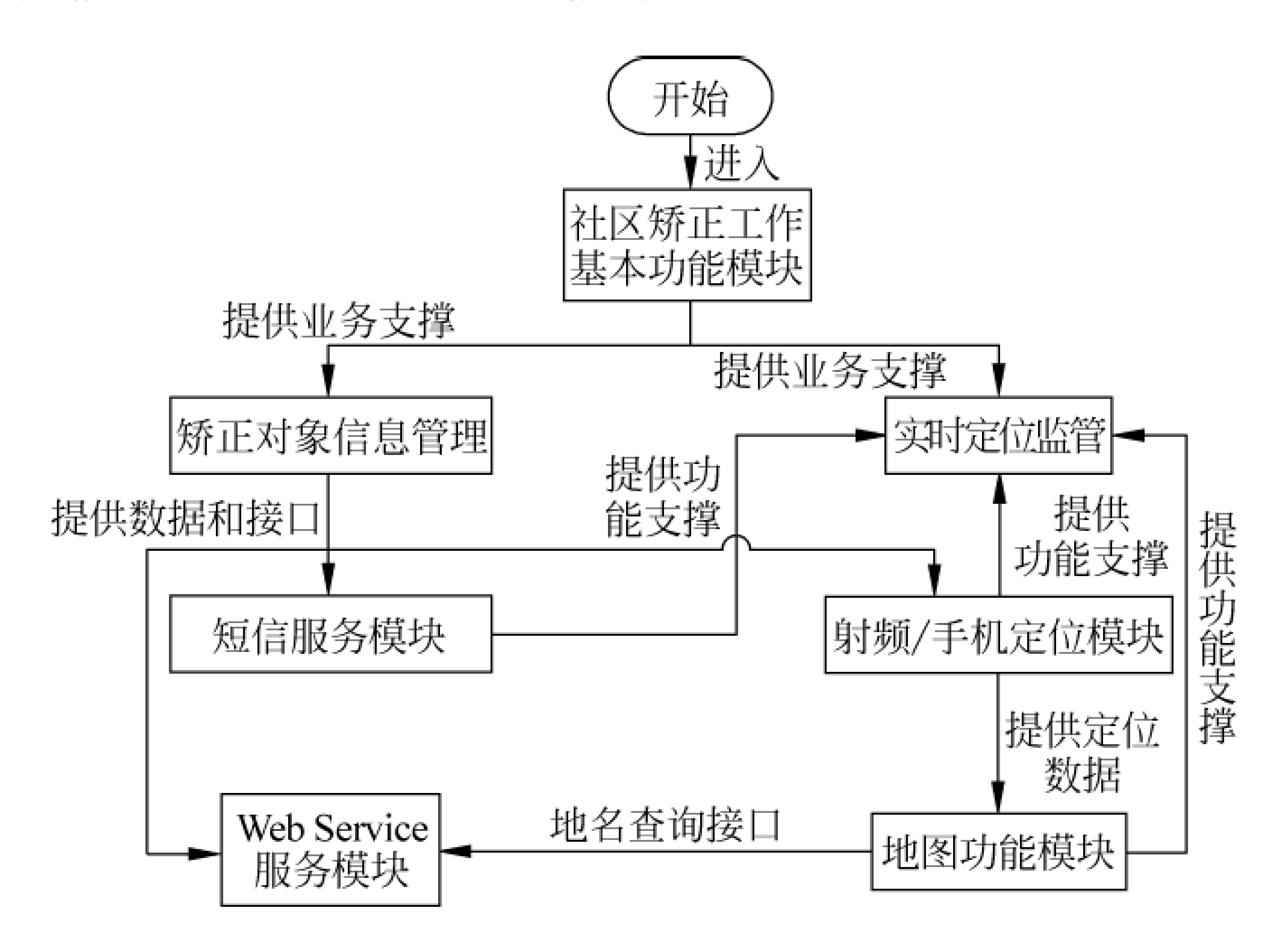


图6-25 位置监控系统基础功能

社区矫正位置监控系统采用手机和电子腕带作为定位终端，在与互联网隔离的业务网内运行，提供与不同移动运营商的通信接口，与运营商的通信定位接口采取加密处理后的安全交换设施进行数据交换。社区矫正涉及矫正信息安全问题及信息的保密性，所以系统需要严格的加密策略和安全防护。

实时定位通过移动终端内嵌的GPS模块获取位置信息，然后将信息发送到服务器端。在Android平台上，系统会根据设备所处环境动态选择GPS定位或基站定位方式进行定位，通过调用getSystemService(Context.LOCATION_SERVICE)取得位置服务，调用requestLocationUpdates(LocationManager.GPS_PROVIDER,1000,1,new LocationListenerImpl())获取位置信息，并将其在窗口显示。由于只能被动地接收GPS卫星广播的数据，因此需要注册一个监听器来处理卫星返回的数据。除了GPS状态，还需监听位置的变化，即定位监听器LocationListener。

位置监听函数是处理GPS位置发生变化时自动回调的方法，我们可从这里获取当前的GPS数据。最后通过回调函数提供的location参数，获取GPS的地理位置信息。

获取地理位置信息包括经纬度、卫星数目、海拔、定位状态等，通过这些信息可在二维地图上标注出该地理位置，供矫正管理员监控。然后，通过二维地图，可将矫正对象的位置信息在地图上实时标记出来，供管理员方便快捷地了解最新的动态。

系统将从终端获取的矫正对象位置信息标注在地图上，这样管理员就可以直观地看到矫正对象的位置信息，要是越界或者违法就可以随时采取措施，同时对矫正对象经常活动地点进行统计分析，可以获取矫正对象的更多信息。

3. 警示告知模块

当矫正对象超出系统设定的安全活动范围时，系统会通过手机自动发送报警，自动备案，并通知警员采取一些措施。警示告知包括越界告警、到期警示、警示处理、警示记录查询等。警示信息需要提示管理员，传统的方法是通过页面实时刷新，然后访问数据库并通知用户有新的消息。但这种方法要不停地刷新服务器，对服务器要求比较高，负担重。Ajax 是异步 JavaScript 及 XML 总称，使用 JavaScript 在 Web 浏览器和 Web 服务器之间发送和接收数据。可以采用 Ajax 实现警示告知功能，通过 JavaScript 和 XML 实现请求服务器对消息提醒，隔段时间检测警示数据库是否有数据，有则根据数据类型弹出提示信息。

1）越界告警

越界告警是限定矫正对象的活动范围，管理员可以根据实际情况在百度地图中添加限定区域，限制矫正对象活动的范围。

2）到期警示

系统通过获取当前时间，对比矫正对象个人资料中的矫正截止日期，判断矫正对象是否到期。到期的矫正人员信息会在到期警示列表显示并提醒管理员办理解矫，若符合解矫条件，则通知矫正对象及时办理解矫手续。

通过表格形式显示所有矫正到期的人员信息，包括矫正人员编号、姓名、管理对象类别、矫正类别、矫正开始日期以及矫正结束日期等。到期警示菜单可以用于查看目前到期的所有人，同时可以进行编辑和删除操作。

3）位置查询

通过矫正人员编号或者矫正人员姓名可以查询矫正对象的位置信息。查询到某个人的位置信息会在位置显示栏显示出来。位置显示主要将矫正对象的位置信息显示到二维地图上，可通过位置查询，查询某个矫正对象并在地图上标记具体位置。

4. 档案管理模块

档案管理模块对整个矫正期内产生的档案资料进行规范化的电子归档整理，按标准一人一档保存在档案服务器中，方便管理员对矫正对象档案的查询和管理；同时建立了矫正对象个人简历，可从矫正对象个人档案中提取出关键信息，以方便管理人员查阅并进行编辑。

档案管理包括添加档案信息、查询详细资料、编辑档案信息以及删除等操作。

5. 信息管理模块

本模块实现发送信息、收件箱、发件箱、草稿箱以及垃圾箱等功能，其业务逻辑流程如图 6-26 所示。管理人员按此流程可操作的步骤如下：

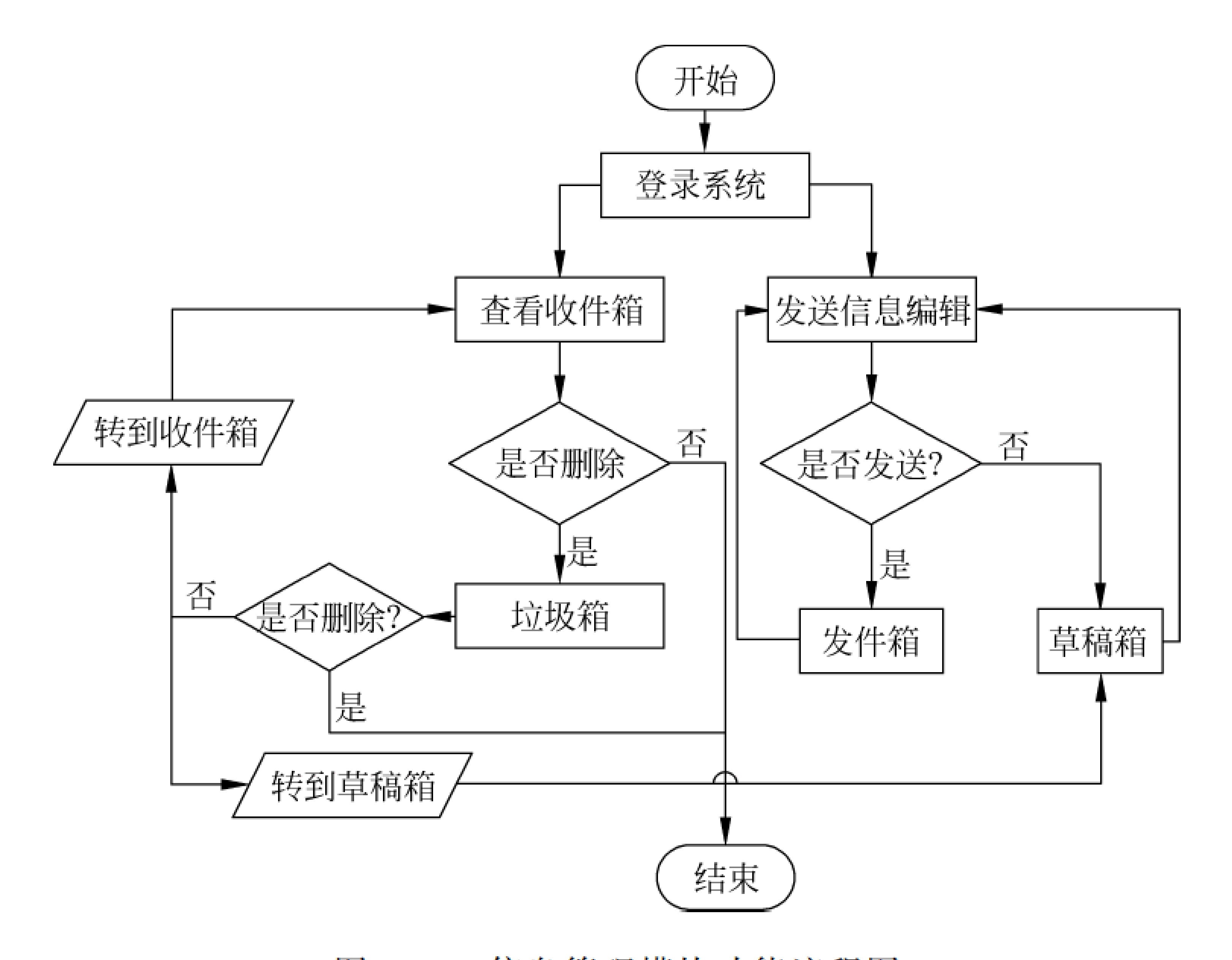


图 6-26　信息管理模块功能流程图

第一，登录系统，通过身份认证登录系统管理。

第二，查看收件箱或者发送信息编辑。

第三，查看发件箱内容是否删除，若删除则会保存在垃圾箱；对于编辑项完成信息单击发送会进入发件箱，不发送则保存在草稿箱。

第四，将垃圾箱中的文件转到收件箱和草稿箱，也可以彻底删除。

参 考 文 献

[1] 朱洪波，杨龙祥，朱琦．物联网技术进展与应用[J]．南京邮电大学学报：自然科学版，2011，31(1)：1-9．

[2] 刘强，崔莉，陈海明．物联网关键技术与应用[J]．计算机科学，2010，37(6)：1-4．

[3] 陈海明，崔莉，谢开斌．物联网体系结构与实现方法的比较研究[J]．计算机学报，2013，36(1)：168-188．

[4] 刘强，崔莉，陈海明．物联网关键技术与应用[J]．计算机科学，2010，37(6)：1-4．

[5] 郭铁．基于 Web Service 的 EPC 信息服务的研究[D]．电子科技大学，2008．

[6] 张文欣．RFID 系统中数据传输技术的分析与研究[D]．合肥工业大学，2012．

[7] 吴虎．EPC 信息服务系统的研究与实现[D]．华中科技大学，2011．

[8] 邓海生．基于 SOA 的 RFID 中间件研究与应用[D]．西安理工大学，2007．

[9] 韩西杰．基于 CEP 的 RFID 数据处理方法研究与实现[D]．西安理工大学，2011．

[10] 刘佳．基于 RFID 技术的监狱管控系统设计与实现[D]．国防科学技术大学，2008．

[11] 郑文斌．远距离射频识别系统结构与模块设计[D]．中南大学，2008．

[12] 赵明华．人脸检测和识别技术的研究[D]．四川大学，2006．

[13] 李永，殷建平，祝恩，胡春风，陈晖．多指纹识别比较研究[J]．计算机工程与科学，2008，12：32-35．

[14] 李程，钱松荣．射频识别动态定位方法[J]．通信学报，2013，(4)：144-148．

[15] 刘军发，谷洋，陈益强等．具有时效机制的增量式无线定位方法[J]．计算机学报，2013，36(7)：1448-1455．

[16] 肖竹，王勇超，田斌，于全，易克初．超宽带定位研究与应用：回顾和展望[J]．电子学报，2011，01：133-141．

[17] 杨震，路保中．社区矫正工作中存在的问题与对策[J]．中国司法，2013，12：86-87．

[18] 王秀玲，李菊英．美国社区矫正制度[C]．首届法律适用国际高层论坛论文集．2005：676-684．

[19] 罗军舟，金嘉晖，宋爱波，东方．云计算：体系架构与关键技术[J]．通信学报，2011，07：3-21．

[20] 张建成，宋丽华，鹿全礼，郭锐，刘永泉．云计算方案分析研究[J]．计算机技术与发展，2012，01：165-167，171．

[21] 陈康，郑纬民．云计算：系统实例与研究现状[J]．软件学报，2009，05：1337-1348．

[22] 范昊，余婷．一种新型的网络分布式计算—云计算[J]．高性能计算技术，2009 (6)：6-10．

[23] 张建勋，古志民，郑超等．云计算研究进展综述[J]．计算机应用研究，2010，27(2)：429-433．

[24] 兰雨晴，申骞，刘铭等．云计算环境中在线迁移技术研究[J]．电信科学，2010，26(9)：90-94．

[25] 孟小峰，慈祥．大数据管理：概念、技术与挑战[J]．计算机研究与发展，2013，50(1)：146-169．

[26] 刘智慧，张泉灵．大数据技术研究综述[J]．浙江大学学报(工学版)，2014，06：957-972．

[27] 刘鹏,吴兆峰,胡谷雨.大数据——正在发生的深刻变革[J].中兴通信技术,2013,04:2-7.
[28] 王秀磊,刘鹏.大数据关键技术[J].中兴通信技术,2013,04:17-21.
[29] 逄利华,张锦春.基于Hadoop的分布式数据库系统[J].办公自动化,2014,05:47-49,
[30] 姜锋.基于Hadoop平台的海量数据处理研究及应用[D].北京邮电大学,2013.7.
[31] 王华婷.社区矫正对象的分类矫治初探[J].北京农学院学报,2008,02:56-59.
[32] 邹复民,蒋新华,胡惠淳,朱铨,庄孝昆.云计算研究与应用现状综述[J].福建工程学院学报,2013,03:231-242.
[33] 白蛟,全春来,郭镇.基于物联网的公共安全云计算平台[J].计算机工程与设计,2011,11:3696-3700.
[34] 李臣杰.基于蜂窝网的手机基站定位算法研究[D].郑州大学,2012.
[35] 梁元诚.基于无线局域网的室内定位技术研究与实现[D].电子科技大学,2009.
[36] 李军怀,孙转宜,王一乐等.基于虚拟参考标签的RFID定位系统构建方法[J].计算机科学,2011,38(4):107-110.
[37] 吕林涛,张亚玲,李军怀等.网络信息安全技术概论(第二版)[M].北京:科学出版社,2010.5.
[38] 国家标准GB/T 9387.2-1995.信息处理系统开放系统互连基本参考模型(第2部分:安全体系结构).中国标准出版社,1995.2.
[39] 张国立,马军.主机安全技术研究[J].信息安全与技术.2010,(6).
[40] 徐晓娟.基于Eucalyptus的云计算安全策略和身份认证系统研发[D].西安理工大学硕士学位论文,2014.
[41] Thomas Erl. SOA Principles Of Service Design[M]. Prentice Hall. 2008.
[42] 曾文英等. ESB原理、构架、实现及应用[J].计算机工程与应用,2008(1):225-228.
[43] Roy Fielding. Architectural Styles And The Design Of Network-Based Software Architectures[M]. University Of California, Irvine. 2000.
[44] Leonard Richardson and Sam Ruby. RESTful Web Services[M]. O'Reilly Media, Inc. 2007.
[45] 丁海防.论中国电信移动业务的信息化应用——以社区矫正信息管理系统为例[J].信息通信,2013,06:250-251.
[46] 朱继团,杨琳,曾蔚.云计算视野中的电子政务协同服务模式[J].电子政务,2013,03:117-121.
[47] 管华丽.我国社区矫正程序构建研究[D].安徽大学,2012.
[48] 王少华.基层(司法所)社区矫正工作若干问题研究[D].苏州大学,2012.
[49] 刘立霞,单福荣.社区矫正协同检察监督研究[J].法学杂志,2014,35(2):120-125.
[50] 高艳华.我国社区矫正风险评估制度的完善[D].湖南大学,2013.
[51] 刘润龙.云计算及关键技术研究[J].数字化用户,2013,06:15-16,40.
[52] 管凯斌.寒亭区司法局社区矫正系统的设计与实现[D].山东大学,2012.
[53] 窦万春,江澄.大数据应用的技术体系及潜在问题[J].中兴通讯技术,2013,19(4):8-16.
[54] 曹润涛.基于Hadoop的移动感知系统的设计与实现[D].西安电子科技大学,2012.
[55] 杨娟.基于云计算的设计服务模式研究及原型应用[D].重庆大学,2012.
[56] 刘宇芳.云计算及其实质的探究[J].惠州学院学报,2010,30(6):48-52.

[57] 张显龙. 云计算安全总体框架与关键技术研究[J]. 信息网络安全，2013，(7)：28-31.
[58] 李连，朱爱红. 云计算安全技术研究综述[J]. 信息安全与技术，2013，4(5)：42-45，52.
[59] 杨斌. 社区矫正管理信息系统研究开发[D]. 西安理工大学，2014.
[60] 贾佩. 基于 Eucalyptus 的云计算应用迁移与部署研究开发[D]. 西安理工大学，2014.
[61] 社区矫正管理信息系统技术规范(SF 05001-2013) [S]. 中华人民共和国司法行业标准，2013.1.
[62] 社区矫正人员定位系统技术规范(SF 05002-2013) [S]. 中华人民共和国司法行业标准，2013.1.